Machine Learning

Лекция #2

Ядерные методы

Ядерные матрицы

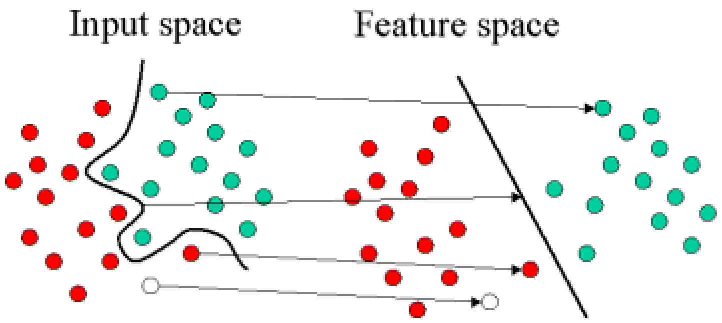
Пусть χ обозначает пространство ввода, которым может быть любой произвольный набор объектов, и пусть $\boldsymbol{D} = [\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_n}] \subset \chi$ обозначает подмножество из n объектов входного пространства. Пусть $\phi: \chi \to \mathcal{F}$ будет отображением из входного пространства в пространство признаков $\mathcal{F}$, которое наделено скалярным произведением и нормой. Пусть $K: \chi \times \chi \to \mathbb{R}$ будет функцией, которая отображает пару входных объектов на значение их скалярного произведения в пространстве признаков, то есть $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$, и пусть $\mathbf{K}$ будет ядерной матрицей $n \times n$, соответствующей подмножеству $\mathbf{D}$.

Функция K называется **положительно полуопределённым ядром** тогда и только тогда, когда она симметрична:

$$K(\mathbf{x}_i, \mathbf{x}_j) = K(\mathbf{x}_j, \mathbf{x}_i)$$

и соответствующая ядерная матрица $\mathbf{K}$ для любого подмножества $\mathbf{D} \subset \chi$ положительно полуопределённая, то есть:

$$\mathbf{a}^T \mathbf{K} \mathbf{a} \geq 0, \text{ for all vectors } \mathbf{a} \in \mathbb{R}^n$$



из чего следует, что:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j K(\mathbf{x}_i, \mathbf{x}_j) \geq 0, \text{ for all } a_i \in \mathbb{R}, i \in [1, n]$$

Сначала удостоверимся, что если $K\left(x_i, x_j\right)$ представляет скалярное произведение $\phi(x_i)^T \phi(x_j)$ в некотором пространстве признаков, то K – положительно полуопределённое ядро. Рассмотрим любой набор данных $\mathbf{D}$, пусть $\mathbf{K} = \{K(\boldsymbol{x_i}, \boldsymbol{x_j})\}$ – соответствующая ядерная матрица. Во-первых, K – симметрично, поскольку скалярное произведение симметрично, из чего следует, что $\mathbf{K}$ – симметрична. Во-вторых, $\mathbf{K}$ – положительно полуопределённая матрица, потому что:

$$\begin{aligned}
\mathbf{a}^T\mathbf{K}\mathbf{a} &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j K(\mathbf{x}_i, \mathbf{x}_j) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) \\
&= \left(\sum_{i=1}^{n} a_i \phi(\mathbf{x}_i)\right)^T \left(\sum_{j=1}^{n} a_j \phi(\mathbf{x}_j)\right) \\
&= \left\| \sum_{i=1}^{n} a_i \phi(\mathbf{x}_i) \right\|^2 \geq 0
\end{aligned}$$

Таким образом, ***K*** является положительным полуопределенным ядром.

Ядерные матрицы

Отображение воспроизводящего ядра

Для отображения воспроизводящего ядра ϕ отобразим каждую точку $\mathbf{x} \in \chi$ в функцию на пространстве функций $[f: \chi \to \mathbb{R}]$, содержащем функции, которые отображают точки из χ в $\mathbb{R}$. Алгебраически это пространство функций представляет собой абстрактное векторное пространство, в котором каждая точка является функцией. В частности, любая точка $\mathbf{x} \in \chi$ во входном пространстве отображается в следующую функцию:

$$\phi(\mathbf{x}) = K(\mathbf{x}, \cdot)$$

где $\cdot$ обозначает любой аргумент в χ. То есть, каждый объект $\mathbf{x}$ во входном пространстве отображается в *точку-функцию* $\phi(\mathbf{x})$, которая, на самом деле, является функцией $K(\mathbf{x},\cdot)$, представляющей её сходство со всеми остальными точками во входном пространстве χ.

Пусть $\mathcal{F}$ – множество всех функций или точек, которые могут быть получены как линейная комбинация любого подмножества точек-функций:

$$\begin{aligned}\mathcal{F} &= span\{K(\mathbf{x}, \cdot) \mid \mathbf{x} \in \mathcal{I}\} \\ &= \left\{\mathbf{f} = f(\cdot) = \sum_{i=1}^{m} \alpha_i K(\mathbf{x}_i, \cdot) \,\middle|\, m \in \mathbb{N}, \alpha_i \in \mathbb{R}, \{\mathbf{x}_1, \ldots, \mathbf{x}_m\} \subseteq \mathcal{I}\right\}\end{aligned}$$

Пусть $\mathbf{f}, \mathbf{g} \in \mathcal{F}$ любые две точки из пространства функций:

$$\mathbf{f} = f(\cdot) = \sum_{i=1}^{m_a} \alpha_i \, K(\mathbf{x}_i, \cdot) \qquad \mathbf{g} = g(\cdot) = \sum_{j=1}^{m_b} \beta_j \, K(\mathbf{x}_j, \cdot)$$

Определим скалярное произведение двух точек как

$$\mathbf{f}^T \mathbf{g} = f(\cdot)^T g(\cdot) = \sum_{i=1}^{m_a} \sum_{j=1}^{m_b} \alpha_i \beta_j K(\mathbf{x}_i, \mathbf{x}_j)$$

Мы можем проверить, что скалярное произведение *билинейно*, то есть линейно по обоим аргументам, потому что

$$\mathbf{f}^T \mathbf{g} = \sum_{i=1}^{m_a} \sum_{j=1}^{m_b} \alpha_i \, \beta_j \, K(\mathbf{x}_i, \mathbf{x}_j) = \sum_{i=1}^{m_a} \alpha_i \, g(\mathbf{x}_i) = \sum_{j=1}^{m_b} \beta_j \, f(\mathbf{x}_j)$$

Из того факта, что **K** - положительное полуопределенное ядро, следует, что

$$\|\mathbf{f}\|^2 = \mathbf{f}^T \mathbf{f} = \sum_{i=1}^{m_a} \sum_{j=1}^{m_a} \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j) \geq 0$$

Таким образом, пространство $\mathcal{F}$ – *предгильбертово пространство*, определяемое как пространство с нормированным внутренним произведением, потому что оно наделено симметричным билинейным скалярным произведением и нормой. Добавляя предельные точки всех сходящихся последовательностей Коши, мы можем превратить $\mathcal{F}$ *в гильбертово пространство*, определяемое как предгильбертово пространство, полное относительно нормы. Однако, рассмотрение данного материала выходит за рамки этой главы.
Пространство $\mathcal{F}$ обладает так называемым *воспроизводящим свойством*, то есть мы можем вычислить функцию $f(\cdot) = \mathbf{f}$ в точке $\mathbf{x} \in \chi$, рассчитав скалярное произведение $\mathbf{f}$ и $\phi(\mathbf{x})$:

$$\mathbf{f}^T\phi(\mathbf{x}) = f(\cdot)^T K(\mathbf{x}, \cdot) = \sum_{i=1}^{m_a} \alpha_i \, K(\mathbf{x}_i, \mathbf{x}) = f(\mathbf{x})$$

По этой причине, пространство $\mathcal{F}$ также называется *гильбертовым пространством с воспроизводящим ядром*. Всё, что нужно сделать – показать, что $K(\mathbf{x}_i, \mathbf{x}_j)$ соответствует скалярному произведению в пространстве признаков $\mathcal{F}$. Это, действительно, так, потому что, используя выражение (5.3), для любых 2 точек-признаков $\phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \in \mathcal{F}$ их скалярное произведение:

$$\phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j) = K(\mathbf{x}_i, \cdot)^T K(\mathbf{x}_j, \cdot) = K(\mathbf{x}_i, \mathbf{x}_j)$$

Отображение Эмпирического Ядра

Отображение воспроизводящего ядра ϕ отображает входное пространство в потенциально бесконечномерное пространство признаков. Однако, при наличии набора данных $\mathbf{D} = \{\mathbf{x}_i\}_{i=1}^n$, мы можем получить конечномерное отображение, вычисляя ядро только в точках в $\mathbf{D}$. То есть определение отображения ϕ выглядит следующим образом:

$$\phi(\mathbf{x}) = \Big(K(\mathbf{x}_1, \mathbf{x}), K(\mathbf{x}_2, \mathbf{x}), \ldots, K(\mathbf{x}_n, \mathbf{x})\Big)^T \in \mathbb{R}^n$$

в котором каждая точка $\mathbf{x} \in \chi$ отображается в n-мерный вектор, содержащий значения ядра $\mathbf{x}$ с каждым из объектов $\mathbf{x}_i \in \mathbf{D}$. Мы можем определить скалярное произведение в пространстве признаков следующим образом:

$$\phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) = \sum_{k=1}^{n} K(\mathbf{x}_k, \mathbf{x}_i) K(\mathbf{x}_k, \mathbf{x}_j) = \mathbf{K}_i^T \mathbf{K}_j$$

где $\mathbf{K}_i$ обозначает i-ый столбец $\mathbf{K}$, который совпадает с i-ой строкой $\mathbf{K}$ (рассматривается как вектор-столбец), так как $\mathbf{K}$ – симметрична. Однако для того, чтобы отображение ϕ было допустимо, мы требуем, чтобы $\phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j)$, поскольку выражение явно этому не удовлетворяет. Единственное решение – замена $\mathbf{K}_i^T \mathbf{K}_j$ в выражении (5.4) на $\mathbf{K}_i^T \mathbf{A} \mathbf{K}_j$ с некоторой положительно полуопределённой матрицей $\mathbf{A}$:

$$\mathbf{K}_i^T \mathbf{A} \mathbf{K}_j = \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j)$$

Если мы можем найти такую матрицу $\mathbf{A}$, это будет означать, что для всех пар отображённых точек мы имеем

$$\left\{\mathbf{K}_i^T \mathbf{A} \mathbf{K}_j\right\}_{i,j=1}^{n} = \left\{K(\mathbf{x}_i, \mathbf{x}_j)\right\}_{i,j=1}^{n}$$

Что может быть записано компактно $$\mathbf{KAK} = \mathbf{K}$$

Это сразу наводит на мысль, что мы берем $\boldsymbol{A} = \boldsymbol{K}^{-\mathbf{1}}$ - (псевдо) обратную матрицу ядра **K**. Модифицированное отображение $\oint$, называемое эмпирическим отображением ядра , определяется как

$$\phi(\mathbf{x}) = \mathbf{K}^{-1/2} \cdot \Big(K(\mathbf{x}_1, \mathbf{x}), K(\mathbf{x}_2, \mathbf{x}), \ldots, K(\mathbf{x}_n, \mathbf{x})\Big)^T \in \mathbb{R}^n$$

так что скалярное произведение д

$$\begin{aligned} \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) &= \Big(\mathbf{K}^{-1/2}\,\mathbf{K}_i\Big)^T \Big(\mathbf{K}^{-1/2}\,\mathbf{K}_j\Big) \\ &= \mathbf{K}_i^T \left(\mathbf{K}^{-1/2}\mathbf{K}^{-1/2}\right) \mathbf{K}_j \\ &= \mathbf{K}_i^T\,\mathbf{K}^{-1}\,\mathbf{K}_j \end{aligned}$$

По всем парам отображенных точек имеем

$$\left\{\mathbf{K}_i^T \mathbf{K}^{-1}\,\mathbf{K}_j\right\}_{i,j=1}^n = \mathbf{K}\,\mathbf{K}^{-1}\,\mathbf{K} = \mathbf{K}$$

Однако важно отметить, что это эмпирическое представление признаков действительно только для *n* точек в **D**.

Ядерные матрицы

Мерсеровское Отображение Ядра

Отображение ядра при конкретных данных

Мерсеровское отображение ядра наиболее понятно после рассмотрения ядерных матриц для набора данных $\mathbf{D}$ во входном пространстве. Поскольку $\mathbf{K}$ – симметричная положительно полуопределённая матрица, она содержит вещественные и неотрицательные собственные значения, и может быть разложена следующим образом:

$$\mathbf{K} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$$

где $\mathbf{U}$ – ортонормированная матрица собственных векторов $\mathbf{u}_i = (u_{i1}, u_{i2}, \dots, u_{in})^T \in \mathbb{R}^n$ (при $i = 1, \dots, n$); $\boldsymbol{\Lambda}$ – диагональная матрица собственных значений; при этом собственные значения обеих матриц ($\mathbf{U}$ и $\boldsymbol{\Lambda}$) расположены в невозрастающем порядке $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$:

$$\mathbf{U} = \begin{pmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \\ | & | & & | \end{pmatrix} \qquad \boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

Ядерная матрица $\mathbf{K}$ может быть записана как спектральная сумма:

$$\mathbf{K} = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \lambda_2 \mathbf{u}_2 \mathbf{u}_2^T + \cdots + \lambda_n \mathbf{u}_n \mathbf{u}_n^T$$

В частности, функция ядра между $\mathbf{x_i}$ и $\mathbf{x_j}$ задаётся как

$$\begin{aligned} \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) &= \lambda_1\, u_{1i}\, u_{1j} + \lambda_2\, u_{2i}\, u_{2j} \cdots + \lambda_n\, u_{ni}\, u_{nj} \\ &= \sum_{k=1}^{n} \lambda_k\, u_{ki}\, u_{kj} \end{aligned}$$

где u_{ki} – означает i-ый элемент собственного вектора $\mathbf{u_k}$. Отсюда следует, что если мы определим отображение Мерсера ϕ следующим образом:

$$\phi(\mathbf{x}_i) = \left(\sqrt{\lambda_1}\, u_{1i}, \sqrt{\lambda_2}\, u_{2i}, \ldots, \sqrt{\lambda_n}\, u_{ni}\right)^T$$

то $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j)$ – это скалярное произведение в пространстве признаков отображённых точек $\phi(\mathbf{x}_i)$ и $\phi(\mathbf{x}_j)$, потому что

$$\begin{aligned}\phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j) &= \left(\sqrt{\lambda_1}\, u_{1i}, \ldots, \sqrt{\lambda_n}\, u_{ni}\right)\left(\sqrt{\lambda_1}\, u_{1j}, \ldots, \sqrt{\lambda_n}\, u_{nj}\right)^T \\ &= \lambda_1\, u_{1i}\, u_{1j} + \cdots + \lambda_n\, u_{ni}\, u_{nj} = K(\mathbf{x}_i, \mathbf{x}_j)\end{aligned}$$

Отметим, что $\mathbf{U}_i = (u_{1i}, u_{2i}, \ldots, u_{ni})^T$ – это i-ая строка $\mathbf{U}$, можем переписать отображение Мерсера ϕ таким образом:

$$\phi(\mathbf{x}_i) = \sqrt{\mathbf{\Lambda}}\mathbf{U}_i$$

Таким образом, значение ядра – это просто скалярное произведение между масштабированными строками $\mathbf{U}$:

$$\phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j) = \left(\sqrt{\mathbf{\Lambda}}\mathbf{U}_i\right)^T\left(\sqrt{\mathbf{\Lambda}}\mathbf{U}_j\right) = \mathbf{U}_i^T\mathbf{\Lambda}\mathbf{U}_j$$

Мерсеровское Отображение Ядра

Для компактных непрерывных пространств, аналогично дискретному случаю в выражении (5.5), значение ядра между любыми двумя точками может быть записано в виде бесконечного спектрального разложения:спектрального разложения

$$K(\mathbf{x}_i, \mathbf{x}_j) = \sum_{k=1}^{\infty} \lambda_k \, \mathbf{u}_k(\mathbf{x}_i) \, \mathbf{u}_k(\mathbf{x}_j)$$

где $[\lambda_1, \lambda_2, \dots]$ – бесконечное множество собственных значений, $\{\mathbf{u_1}(\cdot), \mathbf{u_2}(\cdot), \dots\}$ – соответствующее множество ортогональных и нормализованных *собственных функций*, то есть, каждая функция $\mathbf{u}_{\boldsymbol{i}}(\cdot)$ – решение интегрального уравнения

$$\int K(\mathbf{x}, \mathbf{y}) \, \mathbf{u}_i(\mathbf{y}) \, d\mathbf{y} = \lambda_i \mathbf{u}_i(\mathbf{x})$$

и K – непрерывное положительно полуопределённое ядро, то есть, для всех функций $a(\cdot)$ с конечным квадратным интегралом (например, $\int a(\mathbf{x})^2 d\mathbf{x} < \infty$), K удовлетворяет условию

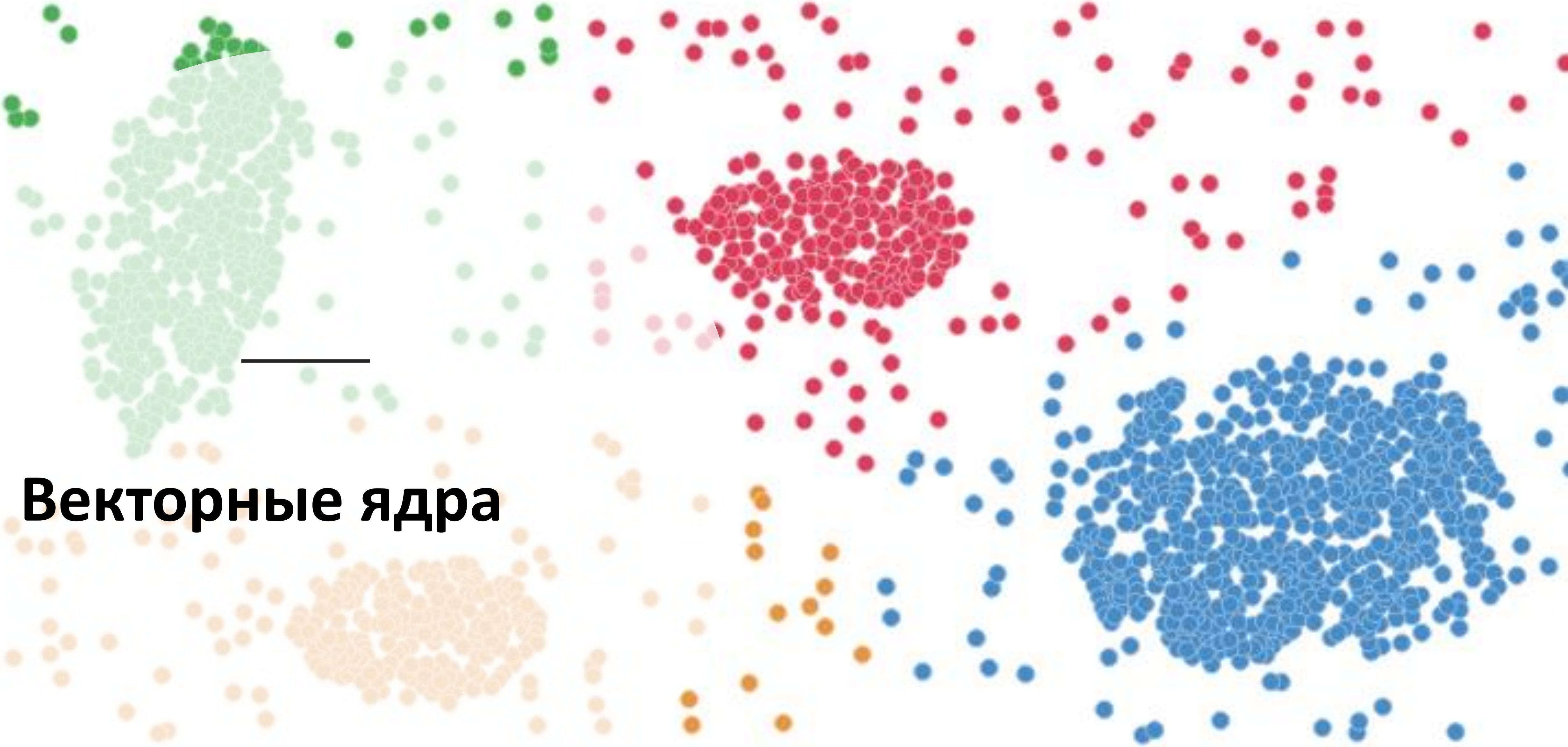
$$\int\int K(\mathbf{x}_1, \mathbf{x}_2) \, a(\mathbf{x}_1) \, a(\mathbf{x}_2) \, d\mathbf{x}_1 \, d\mathbf{x}_2 \geq 0$$

Аналогично отображению Мерсера для конкретных данных, общее отображение ядра Мерсера имеет вид

$$\phi(\mathbf{x}_i) = \left(\sqrt{\lambda_1} \, \mathbf{u}_1(\mathbf{x}_i), \sqrt{\lambda_2} \, \mathbf{u}_2(\mathbf{x}_i), \dots\right)^T$$

со значением ядра, эквивалентным скалярному произведению между двумя сопоставленными точками:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$$



Векторные ядра

Полиномиальное ядро

Полиномиальные ядра бывают двух типов: однородные и неоднородные. Пусть $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$. Однородное полиномиальное ядро определяется как

$$K_q(\mathbf{x}, \mathbf{y}) = \phi(\mathbf{x})^T \phi(\mathbf{y}) = (\mathbf{x}^T \mathbf{y})^q$$

где q - степень многочлена. Это ядро соответствует пространству признаков, охватываемому всеми произведениями ровно q атрибутов. Наиболее типичными случаями являются линейное (q = 1) и квадратичное (q = 2) ядра, заданные как

$$K_1(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{y}$$

$$K_2(\mathbf{x}, \mathbf{y}) = (\mathbf{x}^T \mathbf{y})^2$$

Неоднородное полиномиальное ядро определяется как

$$K_q(\mathbf{x}, \mathbf{y}) = \phi(\mathbf{x})^T \phi(\mathbf{y}) = (c + \mathbf{x}^T \mathbf{y})^q$$

где q - степень многочлена, а c ≥ 0 - некоторая постоянная. При c = 0 получаем однородное ядро. Когда c> 0, это ядро соответствует пространству функций, охватываемому всеми продуктами не более чем с q атрибутами. Это видно из биномиального разложения

$$K_q(\mathbf{x}, \mathbf{y}) = (c + \mathbf{x}^T \mathbf{y})^q = \sum_{k=0}^{q} \binom{q}{k} c^{q-k} \left(\mathbf{x}^T \mathbf{y}\right)^k$$

Например, для типичного значения c = 1 неоднородное ядро представляет собой взвешенную сумму однородных полиномиальных ядер для всех степеней до q, то есть

$$(1+\mathbf{x}^T\mathbf{y})^q = 1 + q\mathbf{x}^T\mathbf{y} + \binom{q}{2}\left(\mathbf{x}^T\mathbf{y}\right)^2 + \cdots + q\left(\mathbf{x}^T\mathbf{y}\right)^{q-1} + \left(\mathbf{x}^T\mathbf{y}\right)^q$$

Для полиномиального ядра можно построить отображение φ из входного пространства в пространство признаков. Пусть $n_0, n_1, \ldots, n_d$ обозначают неотрицательные целые числа, такие что $\sum_{i=1}^{d} n_i = q$. Далее, пусть $n = (n_0, n_1, \ldots, n_d)$, и |n| = $\sum_{i=1}^{d} n_i = q$. Также, пусть $\binom{q}{n}$ обозначает полиномиальный коэффициент

$$\binom{q}{\mathbf{n}} = \binom{q}{n_0, n_1, \ldots, n_d} = \frac{q!}{n_0! n_1! \ldots n_d!}$$

Тогда полиномиальное разложение неоднородного ядра имеет вид

$$\begin{aligned}
K_q(\mathbf{x}, \mathbf{y}) &= (c + \mathbf{x}^T\mathbf{y})^q = \left(c + \sum_{k=1}^{d} x_k y_k\right)^q = (c + x_1 y_1 + \cdots + x_d y_d)^q \\
&= \sum_{|\mathbf{n}|=q} \binom{q}{\mathbf{n}} c^{n_0} (x_1 y_1)^{n_1} (x_2 y_2)^{n_2} \ldots (x_d y_d)^{n_d} \\
&= \sum_{|\mathbf{n}|=q} \binom{q}{\mathbf{n}} c^{n_0} \left(x_1^{n_1} x_2^{n_2} \ldots x_d^{n_d}\right)\left(y_1^{n_1} y_2^{n_2} \ldots y_d^{n_d}\right) \\
&= \sum_{|\mathbf{n}|=q} \left(\sqrt{a_\mathbf{n}} \prod_{k=1}^{d} x_k^{n_k}\right)\left(\sqrt{a_\mathbf{n}} \prod_{k=1}^{d} y_k^{n_k}\right) \\
&= \phi(\mathbf{x})^T \phi(\mathbf{y})
\end{aligned}$$

где $a_n = \binom{q}{n} c^{n_0}$ и суммирование по всем $n = (n_0, n_1, \ldots, n_d)$ таким что $|n| = n_0 + n_1, + \ldots + n_d = q$. Используя нотацию $\boldsymbol{x}^{\boldsymbol{n}} = \prod_{k=1}^{d} x_k^{n_k}$, отображение ϕ: $\mathbb{R}^d \rightarrow \mathbb{R}^m$ представлено в виде вектора

$$\phi(\mathbf{x}) = (\ldots, a_{\mathbf{n}}\mathbf{x}^{\mathbf{n}}, \ldots)^T = \left(\ldots, \sqrt{\binom{q}{\mathbf{n}} c^{n_0}} \prod_{k=1}^{d} x_k^{n_k}, \ \ldots\right)^T$$

Где переменная $n = (n_0, n_1, \ldots, n_d)$ пробегает все возможные присвоения, такие что $|n| = q$. Можно показать, что размерность пространства признаков задается как

$$m = \binom{d+q}{q}$$

Гауссово ядро

Гауссово ядро, также называемое ядром радиальной базисной функции Гаусса (RBF), определяется как

$$K(\mathbf{x}, \mathbf{y}) = \exp\left\{-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}\right\}$$

где σ> 0 - параметр разброса, который играет ту же роль, что и стандартное отклонение в нормальной функции плотности. Обратите внимание, что K (x, x) = 1, и, кроме того, значение ядра обратно пропорционально расстоянию между двумя точками x и y. что пространство признаков для гауссова ядра имеет бесконечную размерность. Чтобы увидеть это, обратите внимание, что экспоненциальная функция может быть записана как бесконечное разложение

$$\exp\{a\} = \sum_{n=0}^{\infty} \frac{a^n}{n!} = 1 + a + \frac{1}{2!}a^2 + \frac{1}{3!}a^3 + \cdots$$

Далее, используя $\gamma = \frac{1}{2\sigma^2}$, и отметив, что $\|x - y\|^2 = \|x\|^2 + \|y\|^2 - 2x^T y$, мы можем переписать Гауссово ядро как:

$$\begin{aligned} K(\mathbf{x}, \mathbf{y}) &= \exp\left\{-\gamma \left\|\mathbf{x} - \mathbf{y}\right\|^2\right\} \\ &= \exp\left\{-\gamma \left\|\mathbf{x}\right\|^2\right\} \cdot \exp\left\{-\gamma \left\|\mathbf{y}\right\|^2\right\} \cdot \exp\left\{2\gamma \mathbf{x}^T \mathbf{y}\right\} \end{aligned}$$

В частности, последний член дается как бесконечное разложение

$$\exp\left\{2\gamma \mathbf{x}^T \mathbf{y}\right\} = \sum_{q=0}^{\infty} \frac{(2\gamma)^q}{q!} \left(\mathbf{x}^T \mathbf{y}\right)^q = 1 + (2\gamma)\mathbf{x}^T \mathbf{y} + \frac{(2\gamma)^2}{2!} \left(\mathbf{x}^T \mathbf{y}\right)^2 + \cdots$$

Используя полиномиальное разложение$(x^T y)^q$, мы можем записать гауссово ядро в виде

$$K(\mathbf{x},\mathbf{y}) = \exp\{-\gamma\|\mathbf{x}\|^2\}\exp\left\{-\gamma\left\|\mathbf{y}\right\|^2\right\}\sum_{q=0}^{\infty}\frac{(2\gamma)^q}{q!}\left(\sum_{|\mathbf{n}|=q}\binom{q}{\mathbf{n}}\prod_{k=1}^{d}(x_k y_k)^{n_k}\right)$$

$$= \sum_{q=0}^{\infty}\sum_{|\mathbf{n}|=q}\left(\sqrt{a_{q,\mathbf{n}}}\exp\{-\gamma\|\mathbf{x}\|^2\}\prod_{k=1}^{d}x_k^{n_k}\right)\left(\sqrt{a_{q,\mathbf{n}}}\exp\left\{-\gamma\left\|\mathbf{y}\right\|^2\right\}\prod_{k=1}^{d}y_k^{n_k}\right)$$

$$= \phi(\mathbf{x})^T\phi(\mathbf{y})$$

где $a_{q,n} = \frac{(2\gamma)^q}{q!}\binom{q}{n}$ и $n = (n_0, n_1, \ldots, n_d)$ с $|n| = n_0 + n_1, + \ldots + n_d = q$ Отображение в пространство признаков соответствует функции ϕ: $\mathbb{R}^d \to \mathbb{R}^\infty$

$$\phi(\mathbf{x}) = \left(\ldots,\ \sqrt{\frac{(2\gamma)^q}{q!}\binom{q}{\mathbf{n}}}\exp\{-\gamma\|\mathbf{x}\|^2\}\prod_{k=1}^{d}x_k^{n_k},\ \ldots\right)^T$$

с размерностями в пределах всех степеней *q = 0, . . . ,∞*, и с переменной $n = (n_1, \ldots, n_d)$ перебирая все возможные присваивания таким образом, что $|n| = q$ для каждого значения q .Так как ϕ отображает входное пространство в бесконечномерное пространство признаков, мы, очевидно, не можем явно преобразовать x в $\phi(x)$, но вычислить Гауссово ядро K(x,y) просто.

Основные операции ядра в пространстве признаков

Норма точки

Норма точки в пространстве признаков вычисляется следующим образом:

$$\|\phi(\mathbf{x})\|^2 = \phi(\mathbf{x})^T\phi(\mathbf{x}) = K(\mathbf{x}, \mathbf{x})$$

что предполагает $\|\phi(\mathbf{x})\| = \sqrt{K(\mathbf{x}, \mathbf{x})}.$

Расстояние между точками

Расстояние между двумя точками $\phi(\boldsymbol{x}_i)$ и $\phi(\boldsymbol{x}_j)$ может быть вычислено как

$$\begin{aligned}\left\|\phi(\mathbf{x}_i) - \phi(\mathbf{x}_j)\right\|^2 &= \|\phi(\mathbf{x}_i)\|^2 + \left\|\phi(\mathbf{x}_j)\right\|^2 - 2\phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j)\\ &= K(\mathbf{x}_i, \mathbf{x}_i) + K(\mathbf{x}_j, \mathbf{x}_j) - 2K(\mathbf{x}_i, \mathbf{x}_j)\end{aligned}$$

что предполагает

$$\delta\big(\phi(\mathbf{x}_i), \phi(\mathbf{x}_j)\big) = \left\|\phi(\mathbf{x}_i) - \phi(\mathbf{x}_j)\right\| = \sqrt{K(\mathbf{x}_i, \mathbf{x}_i) + K(\mathbf{x}_j, \mathbf{x}_j) - 2K(\mathbf{x}_i, \mathbf{x}_j)}$$

Из преобразования уравнения (5.11) видно, что значение ядра можно рассматривать как меру схожести между двумя точками:

$$\frac{1}{2}\left(\|\phi(\mathbf{x}_i)\|^2 + \|\phi(\mathbf{x}_j)\|^2 - \|\phi(\mathbf{x}_i) - \phi(\mathbf{x}_j)\|^2\right) = K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j)$$

Таким образом, чем больше расстояние $\left\|\phi(\boldsymbol{x}_i) - \phi(\boldsymbol{x}_j)\right\|$ между двумя точками в пространстве признаком, тем меньше значение ядра и меньше схожесть.

Среднее в пространстве признаков

Среднее точек в пространстве признаков определено как

$$\boldsymbol{\mu}_\phi = \frac{1}{n}\sum_{i=1}^{n}\phi(\mathbf{x}_i)$$

Поскольку в общем случае к $\phi(\boldsymbol{x}_i)$ доступа нет, нельзя явно вычислить среднюю точку в пространстве признаков. Тем не менее, можно вычислить квадрат нормы среднего следующим образом:

$$\begin{aligned}
\|\boldsymbol{\mu}_\phi\|^2 &= \boldsymbol{\mu}_\phi^T\boldsymbol{\mu}_\phi \\
&= \left(\frac{1}{n}\sum_{i=1}^{n}\phi(\mathbf{x}_i)\right)^T\left(\frac{1}{n}\sum_{j=1}^{n}\phi(\mathbf{x}_j)\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}K(\mathbf{x}_i, \mathbf{x}_j)
\end{aligned}$$

Общая дисперсия в пространстве признаков
Сначала выведем формулу квадрата расстояния от точки $\phi(\boldsymbol{x}_i)$ до среднего $\boldsymbol{\mu}_\phi$ в пространстве признаков.

$$\|\phi(\mathbf{x}_i)-\boldsymbol{\mu}_\phi\|^2 = \|\phi(\mathbf{x}_i)\|^2 - 2\phi(\mathbf{x}_i)^T\boldsymbol{\mu}_\phi + \|\boldsymbol{\mu}_\phi\|^2$$
$$= K(\mathbf{x}_i,\mathbf{x}_i) - \frac{2}{n}\sum_{j=1}^{n} K(\mathbf{x}_i,\mathbf{x}_j) + \frac{1}{n^2}\sum_{a=1}^{n}\sum_{b=1}^{n} K(\mathbf{x}_a,\mathbf{x}_b)$$

Общая дисперсия (ур-е 1.4) в пространстве признаков получается из среднего квадрата отклонения точек от среднего:

$$\sigma_\phi^2 = \frac{1}{n}\sum_{i=1}^{n}\|\phi(\mathbf{x}_i)-\boldsymbol{\mu}_\phi\|^2 = \frac{1}{n}\sum_{i=1}^{n}\left(K(\mathbf{x}_i,\mathbf{x}_i) - \frac{2}{n}\sum_{j=1}^{n} K(\mathbf{x}_i,\mathbf{x}_j) + \frac{1}{n^2}\sum_{a=1}^{n}\sum_{b=1}^{n} K(\mathbf{x}_a,\mathbf{x}_b)\right)$$
$$= \frac{1}{n}\sum_{i=1}^{n} K(\mathbf{x}_i,\mathbf{x}_i) - \frac{2}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} K(\mathbf{x}_i,\mathbf{x}_j) + \frac{n}{n^3}\sum_{a=1}^{n}\sum_{b=1}^{n} K(\mathbf{x}_a,\mathbf{x}_b)$$
$$= \frac{1}{n}\sum_{i=1}^{n} K(\mathbf{x}_i,\mathbf{x}_i) - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} K(\mathbf{x}_i,\mathbf{x}_j)$$

Центрирование в пространстве признаков

Каждую точку в пространстве признаков можно центрировать с помощью вычитания среднего:

$$\hat{\phi}(\mathbf{x}_i) = \phi(\mathbf{x}_i) - \boldsymbol{\mu}_\phi$$

Поскольку в общем случае нет явного представления ни $\phi(\boldsymbol{x}_i)$, ни $\boldsymbol{\mu}_\phi$, нельзя явно центрировать точки. Но можно посчитать *центрированную ядерную матрицу*, т.е. ядерную матрицу на центрированными точками.

Центрированная ядерная матрица задана как

$$\hat{\mathbf{K}} = \left\{ \hat{K}(\mathbf{x}_i, \mathbf{x}_j) \right\}_{i,j=1}^{n}$$

где каждая ячейка соответствует ядру между центрированными точками, т.е.:

$$\begin{aligned}
\hat{K}(\mathbf{x}_i, \mathbf{x}_j) &= \hat{\phi}(\mathbf{x}_i)^T \hat{\phi}(\mathbf{x}_j) \\
&= (\phi(\mathbf{x}_i) - \boldsymbol{\mu}_\phi)^T (\phi(\mathbf{x}_j) - \boldsymbol{\mu}_\phi) \\
&= \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) - \phi(\mathbf{x}_i)^T \boldsymbol{\mu}_\phi - \phi(\mathbf{x}_j)^T \boldsymbol{\mu}_\phi + \boldsymbol{\mu}_\phi^T \boldsymbol{\mu}_\phi
\end{aligned}$$

$$= K(\mathbf{x}_i, \mathbf{x}_j) - \frac{1}{n} \sum_{k=1}^{n} \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_k) - \frac{1}{n} \sum_{k=1}^{n} \phi(\mathbf{x}_j)^T \phi(\mathbf{x}_k) + \|\boldsymbol{\mu}_\phi\|^2$$

$$= K(\mathbf{x}_i, \mathbf{x}_j) - \frac{1}{n} \sum_{k=1}^{n} K(\mathbf{x}_i, \mathbf{x}_k) - \frac{1}{n} \sum_{k=1}^{n} K(\mathbf{x}_j, \mathbf{x}_k) + \frac{1}{n^2} \sum_{a=1}^{n} \sum_{b=1}^{n} K(\mathbf{x}_a, \mathbf{x}_b)$$

Другими словами, можно вычислить центрированную ядерную матрицу с использованием одной только функции ядра. Над всеми парами точек, центрированная ядерная матрица может быть записана следующим образом:

$$\begin{aligned}\hat{\mathbf{K}} &= \mathbf{K} - \frac{1}{n}\mathbf{1}_{n\times n}\mathbf{K} - \frac{1}{n}\mathbf{K}\mathbf{1}_{n\times n} + \frac{1}{n^2}\mathbf{1}_{n\times n}\mathbf{K}\mathbf{1}_{n\times n} \\ &= \left(\mathbf{I} - \frac{1}{n}\mathbf{1}_{n\times n}\right)\mathbf{K}\left(\mathbf{I} - \frac{1}{n}\mathbf{1}_{n\times n}\right)\end{aligned}$$

Где $\mathbf{1}_{n\times n}$ **–** единичная матрица, у которой все элементы равны 1.

Нормализация в пространстве признаков

Частый вид нормализации – убедится, что точки в пространстве признаков имеют единичную длину с помощью замены $\phi(\boldsymbol{x}_i)$ на соответствующий единичный вектор $\phi_n(\boldsymbol{x}_i) = \frac{\phi(\boldsymbol{x}_i)}{\|\phi(\boldsymbol{x}_i)\|}$. Скалярному произведению в пространстве признаков соответствует косинус угла между двумя отображенными точки, поскольку

$$\phi_n(\mathbf{x}_i)^T \phi_n(\mathbf{x}_j) = \frac{\phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)}{\left\|\phi(\mathbf{x}_i)\right\| \cdot \left\|\phi(\mathbf{x}_j)\right\|} = \cos\theta$$

Если отображенные точки и центрированы, и нормализованы, то скалярное произведение – корреляция между двумя точками в пространстве признаков. Нормализованная матрица ядра $\boldsymbol{K}_n$ может быть получена с использованием только функции ядра K следующим образом:

$$\mathbf{K}_n(\mathbf{x}_i, \mathbf{x}_j) = \frac{\phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)}{\left\|\phi(\mathbf{x}_i)\right\| \cdot \left\|\phi(\mathbf{x}_j)\right\|} = \frac{K(\mathbf{x}_i, \mathbf{x}_j)}{\sqrt{K(\mathbf{x}_i, \mathbf{x}_i) \cdot K(\mathbf{x}_j, \mathbf{x}_j)}}$$

Причем у $\boldsymbol{K}_n$ диагональные элементы равны 1.

Пусть $\boldsymbol{W}$ **–** диагональная матрица, состоящая из диагональных элементов $\boldsymbol{K}$

$$\mathbf{W} = \mathrm{diag}(\mathbf{K}) = \begin{pmatrix} K(\mathbf{x}_1, \mathbf{x}_1) & 0 & \cdots & 0 \\ 0 & K(\mathbf{x}_2, \mathbf{x}_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & K(\mathbf{x}_n, \mathbf{x}_n) \end{pmatrix}$$

Нормализованная ядерная матрица в таком случае может быть компактно записана следующим образом:

$$\mathbf{K}_n = \mathbf{W}^{-1/2} \cdot \mathbf{K} \cdot \mathbf{W}^{-1/2}$$

Где $\boldsymbol{W}^{-1/2}$ – диагональная матрица, определенная как $\boldsymbol{W}^{-1/2}(\boldsymbol{x}_i, \boldsymbol{x}_i) = \frac{1}{\sqrt{K(\boldsymbol{x}_i, \boldsymbol{x}_i)}}$, со всеми остальными элементами равными 0.

Ядра для сложных объектов

Ядро спектра для строк

Рассмотрим текст или последовательность данных, определенных над алфавитом Σ. Отображение признаков l-спектра это отображение $\varphi\colon \Sigma^* \to \mathbb{R}^{|\Sigma|^l}$ из множества подстрок над Σ в $|\Sigma|^l$-ное пространство, представляющее собой количество всех возможных подстрок длины l, определяемых как

$$\phi(\mathbf{x}) = \left(\cdots, \#(\alpha), \cdots\right)^T_{\alpha \in \Sigma^l}$$

где $\#(\alpha)$ – количество вхождений строки α длиной l в x. (Полное) отображение спектра является расширением отображения l-спектра, полученного путем рассмотрения всех длин l от 0 до ∞, что приводит к бесконечномерному отображению признаков $\varphi\colon \Sigma \to \mathbb{R}^\infty$:

$$\phi(\mathbf{x}) = \left(\cdots, \#(\alpha), \cdots\right)^T_{\alpha \in \Sigma^*}$$

где $\#(\alpha)$ – количество вхождений строки α в x. Ядро l-спектра между двумя строками $\mathrm{x}_i, \mathrm{x}_j$, — это просто скалярное произведение между их l-отображениями спектра:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$$

Ядра с диффузией на вершинах графа

Пусть $\boldsymbol{S}$ - некоторая симметричная матрица подобия между вершинами графа $G = (V, E)$. Например, $\boldsymbol{S}$ может быть (взвешенной) матрицей смежности $\boldsymbol{A}$ [ур-е. (4.1)] или матрицей Кирхгофа (Лапласа) $\boldsymbol{L} = \boldsymbol{A} - \boldsymbol{\Delta}$ (или ее отрицание), где $\boldsymbol{\Delta}$ - матрица степеней для неориентированного графа G, определенная как $\boldsymbol{\Delta}(i, j) = d_i$ и $\boldsymbol{\Delta}(i, j) = 0$ для всех $i \neq j$ и d_i - степень узла i.

Рассмотрим сходство между любыми двумя вершинами, получаемого путем суммирования произведения сходств по пути длиной 2:

$$S^{(2)}(\mathbf{x}_i, \mathbf{x}_j) = \sum_{a=1}^{n} S(\mathbf{x}_i, \mathbf{x}_a) S(\mathbf{x}_a, \mathbf{x}_j) = \mathbf{S}_i^T \mathbf{S}_j \quad \text{где} \quad \mathbf{S}_i = \Big(S(\mathbf{x}_i, \mathbf{x}_1), S(\mathbf{x}_i, \mathbf{x}_2), \ldots, S(\mathbf{x}_i, \mathbf{x}_n) \Big)^T$$

обозначает вектор (столбец), представляющий i строку $\boldsymbol{S}$ (и, поскольку $\boldsymbol{S}$ симметричен, он также обозначает i столбец $\boldsymbol{S}$). Таким образом, по всем парам узлов матрица подобия по пути длиной 2, обозначенная как $\boldsymbol{S}^{(\mathbf{2})}$, задается как квадрат базовой матрицы подобия $\boldsymbol{S}$:

$$\mathbf{S}^{(2)} = \mathbf{S} \times \mathbf{S} = \mathbf{S}^2$$

В общем случае, если мы просуммируем произведение базовых сходств по всем путям длины l между двумя вершинами, мы получим матрицу подобия $\boldsymbol{S}^l$ длины l, которая является просто l-степенью **S**, то есть,

$$\mathbf{S}^{(l)} = \mathbf{S}^l$$

Степень ядер

Пути четной длины образуют положительно определенные ядра, но пути нечетной длины не гарантируют этого, если только базовая матрица $\boldsymbol{S}$ сама не является положительно определенной. В частности, $\boldsymbol{K} = \boldsymbol{S^2}$ допустимое ядро. Чтобы убедиться в этом, предположим, что i-строка $\boldsymbol{S}$ представляет собой отображение признаков для x_i, то есть $\varphi(\mathrm{x}_i) = S_i$. Значение ядра между любыми двумя точками тогда является скалярным произведением в пространстве признаков.

$$K(\mathbf{x}_i, \mathbf{x}_j) = S^{(2)}(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{S}_i^T \mathbf{S}_j = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$$

Для произвольного пути длины l пусть $\boldsymbol{K} = \boldsymbol{S^l}$. Рассмотрим спектральное разложение матрицы $\boldsymbol{S}$:

$$\mathbf{S} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T = \sum_{i=1}^{n} \mathbf{u}_i \lambda_i \mathbf{u}_i^T$$

где $\boldsymbol{U}$ - ортогональная матрица собственных векторов, а $\boldsymbol{\Lambda}$ - диагональная матрица собственных чисел $\boldsymbol{S}$:

$$\mathbf{U} = \begin{pmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \\ | & | & & | \end{pmatrix} \qquad \boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

Спектральное разложение матрицы $\boldsymbol{K}$ можно получить следующим образом

$$\mathbf{K} = \mathbf{S}^l = \left(\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T\right)^l = \mathbf{U}\left(\boldsymbol{\Lambda}^l\right)\mathbf{U}^T$$

где мы использовали тот факт, что собственные векторы $\boldsymbol{S}$ и $\boldsymbol{S}^l$ идентичны, и, кроме того, собственные числа $\boldsymbol{S}^l$ заданы как $(\lambda_i)^l$ (для всех $\mathrm{i} = 1, \dots, \mathrm{n}$), где λ_i - собственное число $\boldsymbol{S}$. Чтобы $\boldsymbol{K} = \boldsymbol{S}^l$ была положительной определенной матрицей, все ее собственные числа должны быть неотрицательными, что гарантируется для всех путей четной длин. Поскольку $(\lambda_i)^l$ будет отрицательным, если l нечетно при отрицательном λ_i, пути нечетной длины приводят к положительно определенному ядру, только если $\boldsymbol{S}$ положительно определено.

Ядро с экспоненциальной диффузией

Вместо того, чтобы фиксировать длину пути априори, мы можем получить новое ядро между вершинами графа, рассматривая обходы всех возможных длин, при этом уменьшая вклад более длинных путей, которые приводят к ядру с экспоненциальной диффузией, определяемому как

$$\mathbf{K} = \sum_{l=0}^{\infty} \frac{1}{l!} \beta^l \mathbf{S}^l$$

$$= \mathbf{I} + \beta\mathbf{S} + \frac{1}{2!}\beta^2\mathbf{S}^2 + \frac{1}{3!}\beta^3\mathbf{S}^3 + \cdots$$

$$= \exp\{\beta\mathbf{S}\}$$

где β - коэффициент затухания, а $exp\{\beta\boldsymbol{S}\}$ – экспонента матрицы. Ряд в правой части выше сходится при всех β ≥ 0.

Подставив $\boldsymbol{S} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \lambda_i \boldsymbol{u_i}\, \boldsymbol{u_i}^T$ в ур-е (5.15) и используя тот факт, что $\boldsymbol{U}\boldsymbol{U}^T = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{u_i}\, \boldsymbol{u_i}^T = \boldsymbol{I}$, имеем

$$
\begin{aligned}
\mathbf{K} &= \mathbf{I} + \beta\mathbf{S} + \frac{1}{2!}\beta^2\mathbf{S}^2 + \cdots \\
&= \left(\sum_{i=1}^{n} \mathbf{u}_i \mathbf{u}_i^T\right) + \left(\sum_{i=1}^{n} \mathbf{u}_i \beta\lambda_i \mathbf{u}_i^T\right) + \left(\sum_{i=1}^{n} \mathbf{u}_i \frac{1}{2!}\beta^2\lambda_i^2 \mathbf{u}_i^T\right) + \cdots \\
&= \sum_{i=1}^{n} \mathbf{u}_i \left(1 + \beta\lambda_i + \frac{1}{2!}\beta^2\lambda_i^2 + \cdots\right)\mathbf{u}_i^T \\
&= \sum_{i=1}^{n} \mathbf{u}_i \exp\{\beta\lambda_i\}\, \mathbf{u}_i^T \\
&= \mathbf{U} \begin{pmatrix} \exp\{\beta\lambda_1\} & 0 & \cdots & 0 \\ 0 & \exp\{\beta\lambda_2\} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \exp\{\beta\lambda_n\} \end{pmatrix} \mathbf{U}^T
\end{aligned}
$$

Таким образом, собственные векторы $\boldsymbol{K}$ такие же, как и у $\boldsymbol{S}$, тогда как ее собственные числа задаются как $\exp\{\beta\lambda_i\}$, где λ_i - собственное число $\boldsymbol{S}$. Кроме того, $\boldsymbol{K}$ симметричная матрица, потому что $\boldsymbol{S}$ симметрична, а его собственные числа действительны и неотрицательны, потому что результат взятия экспоненты от действительного числа неотрицателен. Таким образом, $\boldsymbol{K}$ является положительной определенной матрицей ядра. Сложность вычисления ядра с диффузией составляет $O(n^3)$, что соответствует сложности спектрального разложение матрицы.

Ядро с диффузией фон Неймана

Связанное ядро, на основе степеней $\boldsymbol{S}$, является ядром с диффузией фон Неймана, определяемое как

$$\mathbf{K} = \sum_{l=0}^{\infty} \beta^l \mathbf{S}^l$$

где $\beta \geq 0$. Расширим уравнение:

$$\begin{aligned}\mathbf{K} &= \mathbf{I} + \beta\mathbf{S} + \beta^2\mathbf{S}^2 + \beta^3\mathbf{S}^3 + \cdots \\ &= \mathbf{I} + \beta\mathbf{S}(\mathbf{I} + \beta\mathbf{S} + \beta^2\mathbf{S}^2 + \cdots) \\ &= \mathbf{I} + \beta\mathbf{S}\mathbf{K}\end{aligned}$$

Преобразуя члены в предыдущем уравнении, мы получаем выражение в замкнутой форме для ядра фон Неймана

$$\mathbf{K} - \beta\mathbf{S}\mathbf{K} = \mathbf{I}$$

$$(\mathbf{I} - \beta\mathbf{S})\mathbf{K} = \mathbf{I}$$

$$\mathbf{K} = (\mathbf{I} - \beta\mathbf{S})^{-1}$$

Подставляя спектральное разложение $\boldsymbol{S} = \boldsymbol{U\Lambda U}^T$ и переписывая $\boldsymbol{I} = \boldsymbol{UU}^T$, мы имеем

$$\begin{aligned}\mathbf{K} &= \left(\mathbf{U}\mathbf{U}^T - \mathbf{U}(\beta\boldsymbol{\Lambda})\mathbf{U}^T\right)^{-1} \\ &= \left(\mathbf{U}\left(\mathbf{I} - \beta\boldsymbol{\Lambda}\right)\mathbf{U}^T\right)^{-1} \\ &= \mathbf{U}\left(\mathbf{I} - \beta\boldsymbol{\Lambda}\right)^{-1}\mathbf{U}^T\end{aligned}$$

где $(\boldsymbol{I} - \beta\boldsymbol{\Lambda})^{-1}$ - диагональная матрица, i-диагональный элемент которой равен $(1 - \beta\lambda_i)^{-1}$. Собственные вектора $\boldsymbol{K}$ и $\boldsymbol{S}$ идентичны, но собственные числа $\boldsymbol{K}$ заданы как $1\ /\ (1 - \beta\lambda_i)$. Чтобы $\boldsymbol{K}$ было положительным определенным ядром, все его собственные числа должны быть неотрицательными, что, в свою очередь, означает, что

$$\begin{aligned}(1 - \beta\lambda_i)^{-1} &\geq 0 \\ 1 - \beta\lambda_i &\geq 0 \\ \beta &\leq 1/\lambda_i\end{aligned}$$

Далее, обратная матрица$(\boldsymbol{I} - \beta\boldsymbol{\Lambda})^{-1}$ существует, только если

$$\det(\mathbf{I} - \beta\boldsymbol{\Lambda}) = \prod_{i=1}^{n}(1 - \beta\lambda_i) \neq 0$$

откуда следует, что $\beta \neq 1/\lambda_i$ для всех i. Таким образом, чтобы $\boldsymbol{K}$ было допустимым ядром, мы требуем, чтобы $\leq 1/\lambda_i$ для всех $\mathrm{i} = 1, \dots, \mathrm{n}$. Таким образом, ядро фон Неймана будет положительно определенным, если $|\beta| \leq 1/\mathrm{p}(\boldsymbol{S})$, где $\mathrm{p}(\boldsymbol{S}) = \max_i\{|\lambda_i|\}$, называемым спектральным радиусом $\boldsymbol{S}$ и определяемым как наибольшее собственное число $\boldsymbol{S}$ по абсолютной величине.

Вопросы?