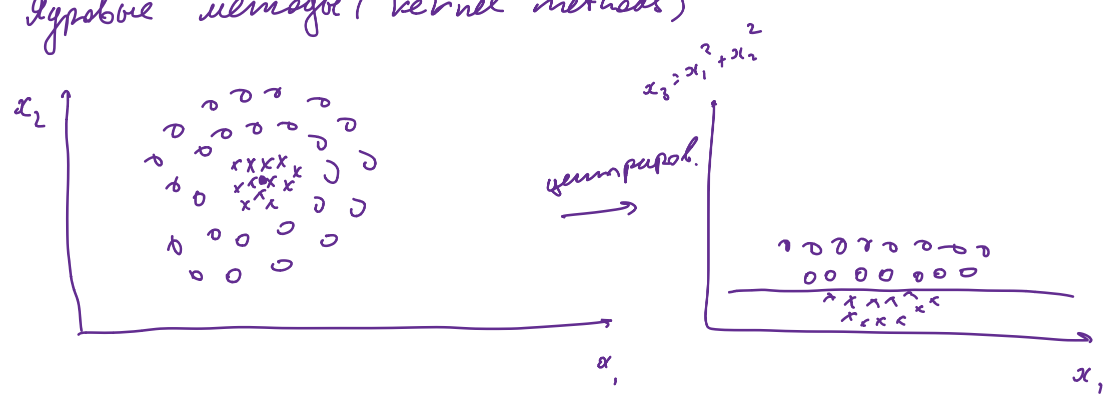
Итог = Округление(0.4 · ДЗ + 0.1 · ПР + 0.2 · К + 0.3 · Э)

Проект

Темы:

1) ядровые методы
2) вероятностные методы
3) кластеризация
   одноклассовые методы
   частичное обучение
4) обучение с учителем
   - метрич. методы
   - быстрый поиск соседей
   - ранжирование
   - рек. системы
5) разное

Ядровые методы (kernel methods)



$x = (x_1, \dots x_d)$

базисные функции:

$\varphi_1(x), \dots \varphi_m(x)$

$$a(x) = \sum_{j=1}^{m} w_j \varphi_j(x)$$

для хорошего качества надо
много базисных ф-ций

(1) Двойственное представление для лин. рег.

$$Q(w) = \frac{1}{2} \sum_{i=1}^{\ell} \left(\sum_{j=1}^{m} w_j \varphi_j(x_i) - y_i \right)^2 + \frac{\lambda}{2} \|w\|_2^2 =$$

$$= \frac{1}{2} \|\Phi w - y\|_2^2 + \frac{\lambda}{2} \|w\|_2^2 \to \min_w$$

$$\Phi = \begin{pmatrix} \varphi_1(x_1) & \varphi_2(x_1) & \dots & \varphi_m(x_1) \\ \dots & & & \\ \varphi_1(x_\ell) & \varphi_2(x_\ell) & \dots & \varphi_m(x_\ell) \end{pmatrix}$$

$$\nabla_w Q = \Phi^T(\Phi w - y) + \lambda w = 0$$

$$w = -\frac{1}{\lambda} \Phi^T (\Phi w - y) = \Phi^T \underbrace{\left(-\frac{1}{\lambda} (\Phi w - y) \right)}_{a}$$

$w = \Phi^T a$ – решение можно искать в таком виде

$$Q(\alpha) = \frac{1}{2}\|\underbrace{\Phi\Phi^T}\alpha - y\| + \frac{\lambda}{2}\alpha^T \underbrace{\Phi\Phi^T}\alpha \to \min_\alpha$$

$\Phi\Phi^T$ – матрица Грама (матрица попарных скалярных произв. объектов)

$$(\Phi\Phi^T)_{ij} = \langle \varphi(x_i), \varphi(x_j) \rangle$$

$$\varphi(x_i) = (\varphi_1(x_i), \dots \varphi_m(x_i))$$

можно записать $Q(w)$ так, что он зависит только от скалярных произв. объектов

② SVM

$$\begin{cases} \sum_{i=1}^{\ell} \lambda_i - \frac{1}{2}\sum_{i,j=1}^{\ell} \lambda_i \lambda_j y_i y_j \langle x_i, x_j \rangle \to \max_\lambda \\ 0 \le \lambda_i \le C \\ \sum_{i=1}^{\ell} \lambda_i y_i = 0 \end{cases}$$

③ Пример добавления признаков

$x_1, \dots x_d$ – исходные признаки

новые признаки:

$$\varphi(x) = (x_i x_j)_{i,j=1}^{d} \in \mathbb{R}^{d^2}$$

$$x, z: \langle \varphi(x), \varphi(z) \rangle = \sum_{i,j=1}^{d} \underbrace{x_i x_j}\, \underbrace{z_i z_j} =$$

$$= \sum_{i=1}^{d}\sum_{j=1}^{d} x_i x_j z_i z_j = \underbrace{\sum_{i=1}^{d} x_i z_i}\, \underbrace{\sum_{j=1}^{d} x_j z_j} = \langle x, z \rangle^2$$

1) добавляем новые признаки $\varphi(x)$

2) по том, что $\langle \varphi(x), \varphi(z) \rangle$ легко выражается через $\langle x, z \rangle$

3) пользуемся методом, который зависит только от $\langle x, z \rangle$

4) $\langle x, z \rangle \to \langle \varphi(x), \varphi(z) \rangle$

5) PROFIT

kernel trick

④ Ядра

Ядро - функция $K(x,z) = \langle \underline{\varphi(x)}, \underline{\varphi(z)} \rangle$,

где $\varphi: \mathbb{X} \to H$

H - спрямляющее пр-во

φ - спрямляющее отображение

Что мы делаем: заменяем в функционале ошибки $\langle x, z \rangle$ на $K(x,z)$

В чем проблема: непонятно, откуда взять $K(x,z)$

Как строить ядра?

$K(x,z)$ - как понять, что это ядро?

Теорема Мерсера.

$K(x,z)$ - ядро $\Leftrightarrow$ $\begin{cases} 1) \; K(x,z) = K(z,x) \\ 2) \; K \text{ неотриц. опред.} \end{cases}$

$\forall \ell \;\; \forall (x_1, \dots x_\ell) \subset \mathbb{R}^d$

$\left(K(x_i, x_j)\right)_{i,j=1}^{\ell}$ неотриц. опред.

$K(x,z) = x z^2$ - не ядро

Теорема 1.

$K_1(x,z)$, $K_2(x,z)$ - ядра; $x, z \in X$

$f(x)$ - веществ. ф-ция на X

$\varphi : X \to \mathbb{R}^n$

K_3 - ядро, заданное на $\mathbb{R}^n$

Тогда след. ф-ции являются ядрами:

1) $K(x,z) = K_1(x,z) + K_2(x,z)$

2) $K(x,z) = \alpha K_1(x,z)$, $\alpha > 0$, $\alpha \in \mathbb{R}$

3) $K(x,z) = K_1(x,z) K_2(x,z)$

4) $K(x,z) = f(x) f(z)$

5) $K(x,z) = K_3(\varphi(x), \varphi(z))$

Теорема 2. $K_1(x,z), K_2(x,z), \dots$ - послед. ядер

и $\exists K(x,z) = \lim\limits_{n \to \infty} K_n(x,z) \quad \forall x, z$

Тогда K - ядро

⑤ Полиномиальные ядра

$p(v)$ - многочлен с неотриц. коэф.

$K(x,z) = p(\langle x,z \rangle)$ - ядро

$$K(x,z) = w_0 + w_1 \langle x,z \rangle + w_2 \langle x,z \rangle^2 + \ldots$$

$\langle x,z \rangle^p$ - ядро по п.3

$w_p \langle x,z \rangle^p$ - ядро по п.2

сумма - ядро по п.1

$$K_m(x,z) = (\langle x,z \rangle + R)^m, \quad R \in \mathbb{R}, \; R > 0$$

$$= \sum_{i=0}^{m} \underbrace{C_m^i R^{m-i}}_{w_i} \underbrace{\langle x,z \rangle^i}$$

если расписать $\langle x,z \rangle^i$, то получим все мономы степени i от исх. признаков

т.е. K_m строит мономы всех степеней до m

моном над $(x_1, \ldots x_d)$ степени k

$$x_{i_1} x_{i_2} \ldots x_{i_k}$$

Зачем нужен R?

коэф. при мономе степени i: $\sqrt{C_m^i R^{m-i}}$

сравним веса при степенях 1 и $(m-1)$:

$$\sqrt{\frac{\cancel{C_m^{m-1}}\, R}{\cancel{C_m^{1}}\, R^{m-1}}} = \sqrt{\frac{1}{R^{m-2}}}$$

R большое, то мономы высоких степеней имеют низкий вклад