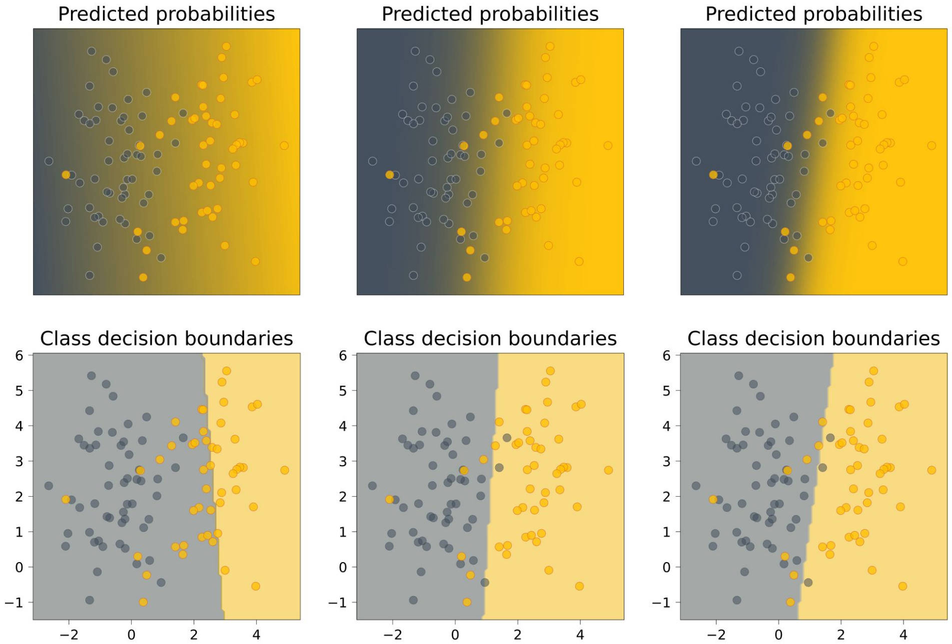
Введение в АД

Лекция 4
SVM, PCA, SVD

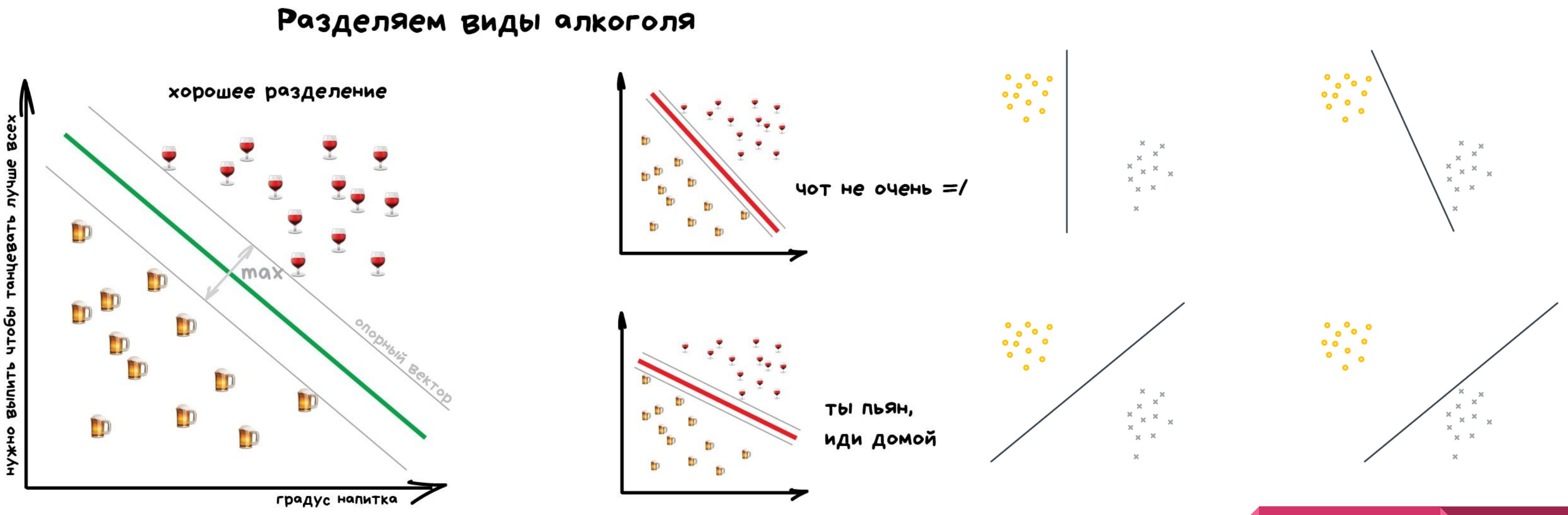
ФЭФМ МФТИ
Весенний семестр 2023

На подумать

В каком классификаторе регуляризационный параметр больше?



Проблема – какой классификатор выбрать?



SVM – support vector machine (метод опорных векторов)

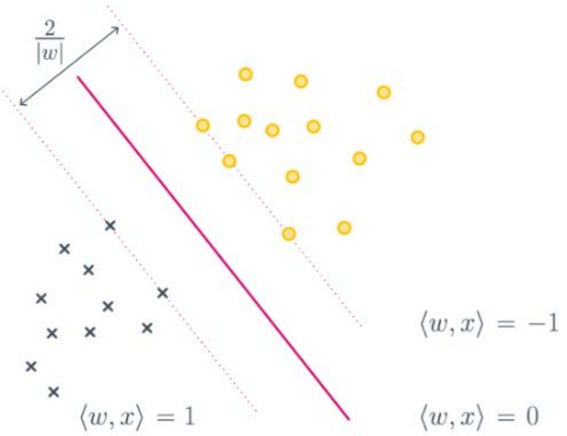
Хотим максимизировать “зазор” между классами

Это можно сделать, слегка поменяв функцию ошибки, а именно положив её равной:

$$F(M) = \max(0, 1 - M)$$

$$L(w, x, y) = \lambda||w||_2^2 + \sum_i \max(0, 1 - y_i\langle w, x_i\rangle)$$

$$\nabla_w L(w, x, y) = 2\lambda w + \sum_i \begin{cases} 0, & 1 - y_i\langle w, x_i\rangle \leq 0 \\ -y_i x_i, & 1 - y_i\langle w, x_i\rangle > 0 \end{cases}$$



Откуда это взялось?

$$\lambda|w|_2^2 + \sum_i \max(0, 1 - y_i\langle w, x_i\rangle) \longrightarrow \min_w$$

Первое слагаемое – максимизируем 2/|w|, второе – штрафуем за ошибку

Нелинейный SVM

Написанное ранее эквивалентно двойственной задаче оптимизации (через условие Лагранжа):

$$\begin{cases} w = \sum_{i=1}^{l} \lambda_i y_i x_i; \\ w_0 = < w, x_i > -y_i, i : \lambda_i > 0; M_i = 1. \end{cases}$$

Можем получить общее выражение для SVM классификатора

$$a(x) = \operatorname{sign}\left(\sum_{i=1}^{\ell} \lambda_i y_i \langle x, x_i \rangle - w_0 \right)$$

А если взять что-то другое как меру “похожести”?

Kernel trick

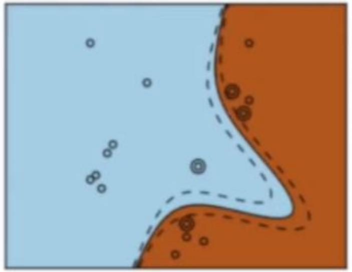
- <u>**Kernel function**</u>

$$K(x, x') : X \times X \to R$$
$$\exists \psi : X \to H : K(x, x') = <\psi(x), \psi(x')>, \text{ H - Hilbert space}$$

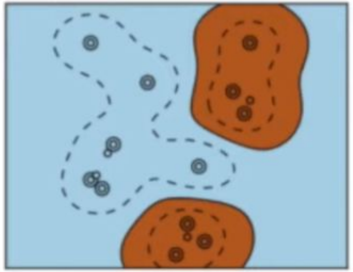
- Kernel examples

$$K(x, x') = <x, x'>^2 \text{ - quadratic}$$
$$K(x, x') = <x, x'>^d \text{ - polynomial with degree } d$$
$$K(x, x') = (<x, x'> + 1)^d \text{ - polynomial with degree } \leq d$$
$$K(x, x') = exp(-\gamma ||x - x'||^2) \text{ - Radial Basis Functions (RBF) kernel}$$

$\left(\langle x, x' \rangle + 1\right)^d, \quad d{=}3$



$\exp\left(-\gamma \|x - x'\|^2\right)$



H еще называют спрямляющим пространством

PCA – Principal component analysis (метод главных компонент)

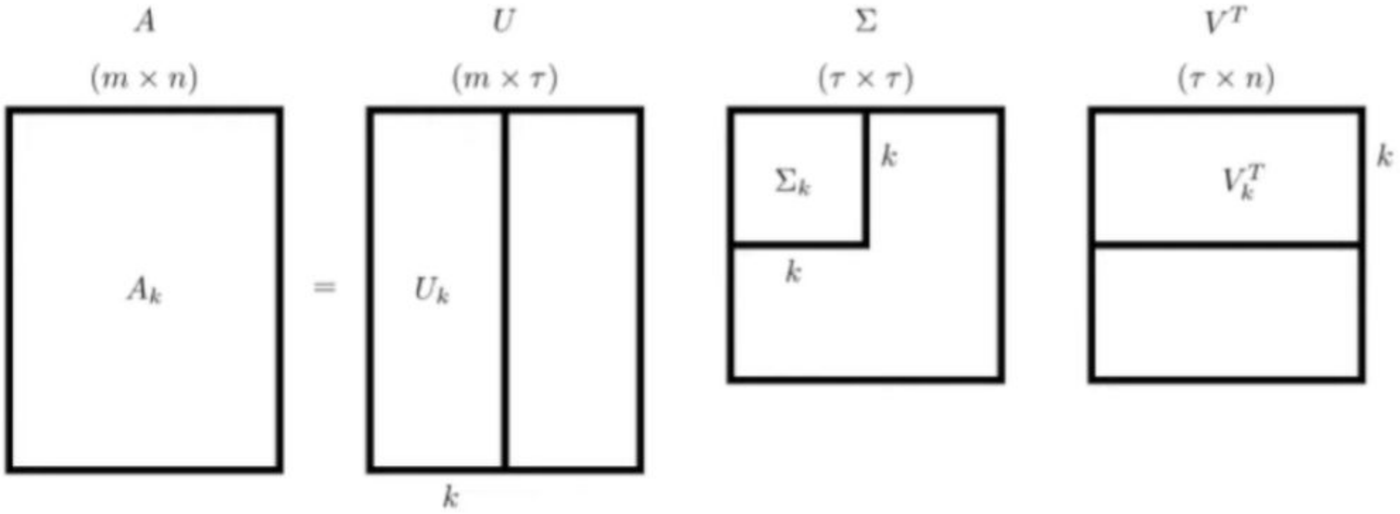
Хотим уменьшить размерность задачи без потери информации

$X_{l,d} \approx U_{l,k} \cdot V_{k,d}^T$ $\qquad ||X - UV^T|| \to min$

Сингулярное разложение! (SVD)

$$A = U\Sigma V^T$$
$$A_k = U_k \Sigma_k V_k^T = (U_k \Sigma_k) V_k^T = U_k (\Sigma_k V_k^T)$$

A_k – Приближение матрицы A матрицей ранга k



Теорема Экарта-Янга (Eckart-Young) – данное приближение является наилучшим

$$X = U\Sigma V^T$$

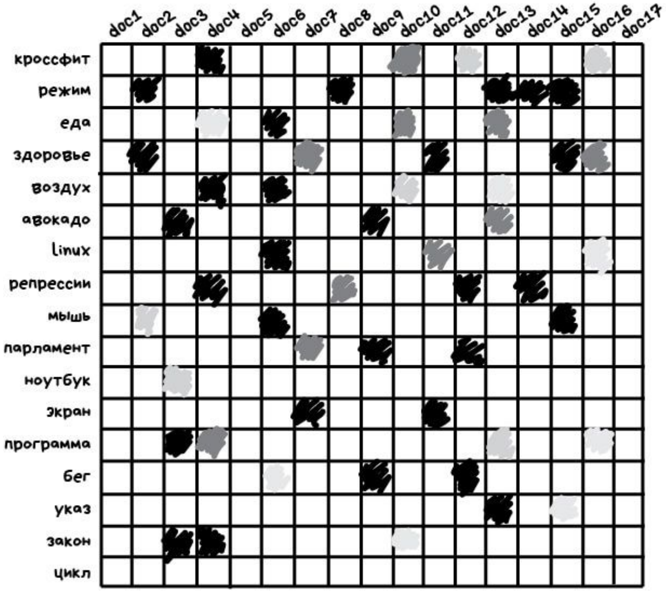
orthogonal

orthogonal

diagonal: sigmas ~ variance

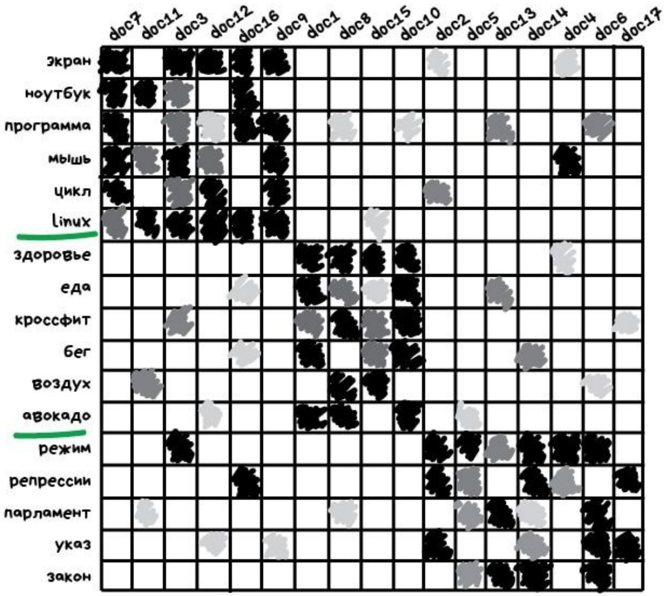
- Consider columns of matrix V new basis vectors: **principal directions**
- Columns of matrix US are called **principal components** of the data
- Singular values are sorted: truncated SVD gives the best projection of dim K

Разделение документов по темам



SVD

2. Раскладываем



1. Строим матрицу как часто каждое слово встречается в каждом документе (чернее - чаще)

3. Получаем наглядные кластера по тематикам (даже если слова не встречались вместе)

Латентно-семантический Анализ (LSA)