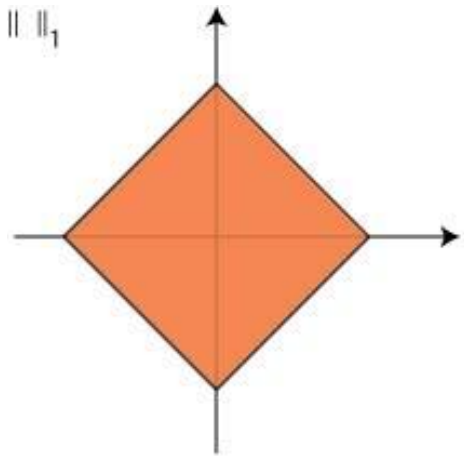
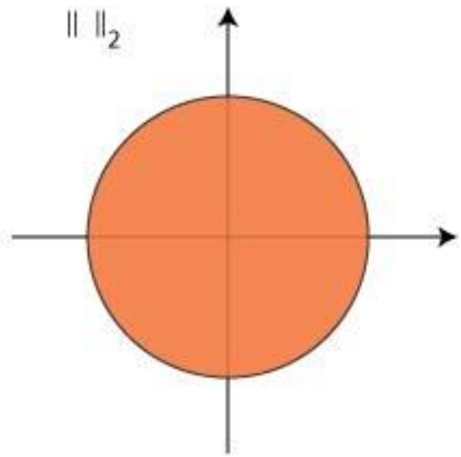
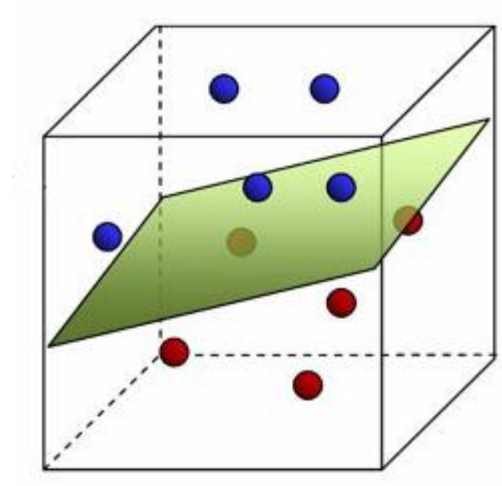






Анализ данных на практике

Линейная классификация

Виктор Кантор

Линейная классификация

- Идея линейной классификации
- Функции потерь
- Градиентный спуск и стохастический градиент
- Регуляризация
- Стандартные линейные классификаторы

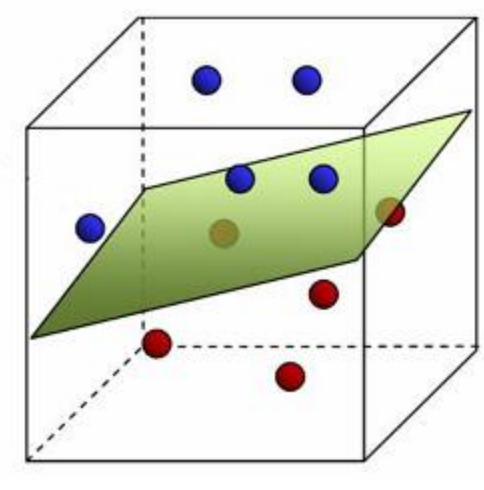
Идея линейной классификации

$$a(x) = sign\{f(x)\}$$

$f(x)$ – дискриминантная функция

$$f(x) = < w, x > + w_0$$

Геометрическая интерпретация: разделяющая гиперплоскость



Отступ (margin)

Отступом алгоритма $a(x) = sign\{f(x)\}$ на объекте x_i называется величина $M_i = y_i f(x_i)$ (y_i - класс, к которому относится x_i)

$$M_i \leq 0 \Leftrightarrow y_i \neq a(x_i)$$
$$M_i > 0 \Leftrightarrow y_i = a(x_i)$$

Функция потерь

$$Q(w) = \sum_{i=1}^{\ell} \left[M_i(w) < 0 \right]$$

Функция потерь

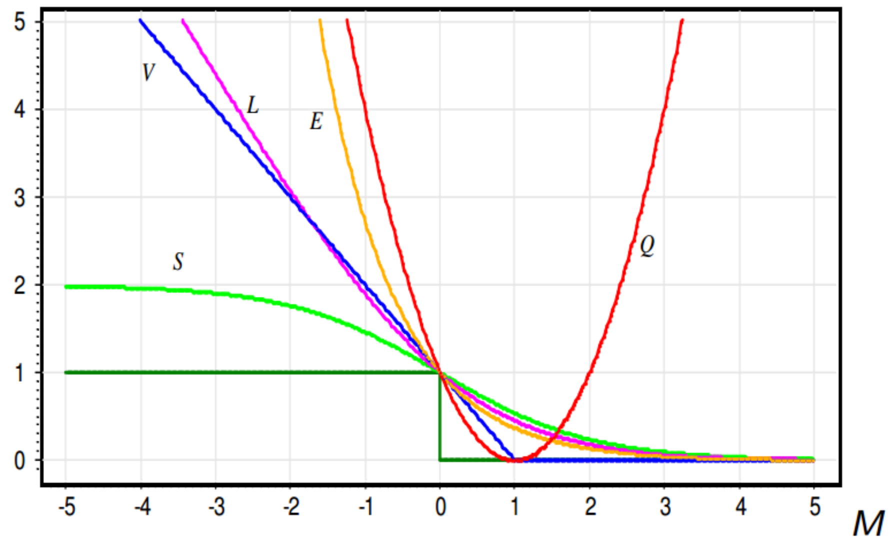
$$Q(w) = \sum_{i=1}^{\ell} \left[M_i(w) < 0\right] \leqslant \widetilde{Q}(w) = \sum_{i=1}^{\ell} \mathscr{L}\left(M_i(w)\right) \to \min_w;$$

Функция эмпирического риска

Функция потерь

Функция потерь

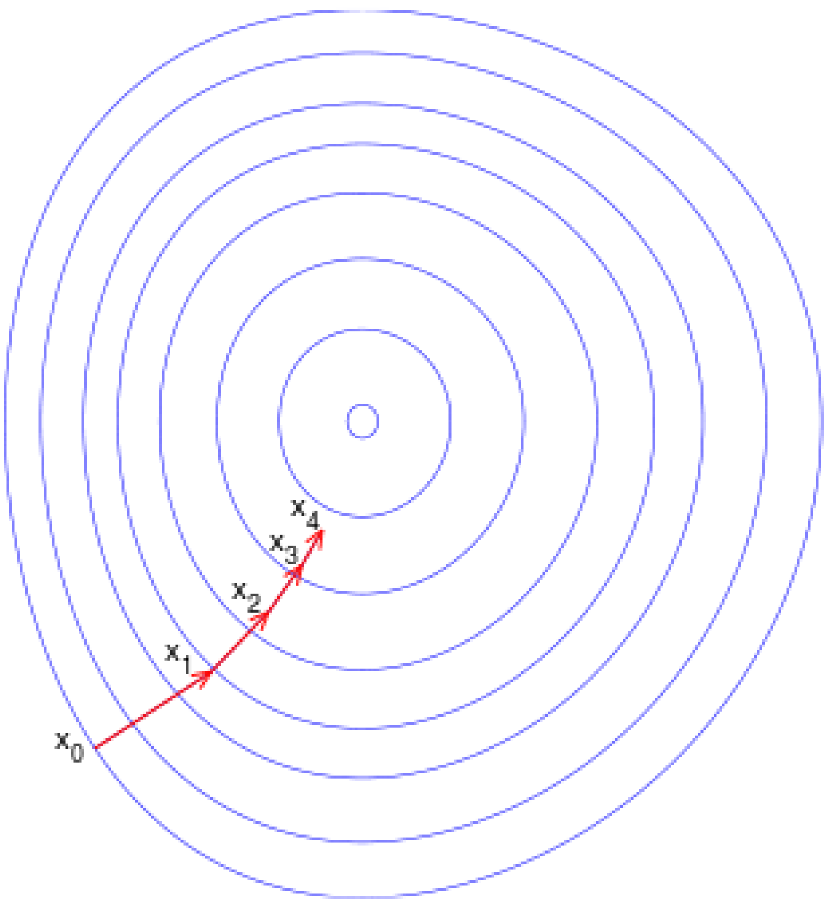
$$Q(w) = \sum_{i=1}^{\ell} [M_i(w) < 0] \leqslant \widetilde{Q}(w) = \sum_{i=1}^{\ell} \mathscr{L}(M_i(w)) \to \min_w;$$



$$Q(M) = (1-M)^2$$
$$V(M) = (1-M)_+$$
$$S(M) = 2(1+e^M)^{-1}$$
$$L(M) = \log_2(1+e^{-M})$$
$$E(M) = e^{-M}$$

Градиентный спуск

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \gamma_n \nabla F(\mathbf{x}_n),\ n \geq 0.$$



$$\nabla_w \tilde{Q} = \sum_{i=1}^{l} \nabla L(M_i)$$

$$\nabla \tilde{Q} = \sum_{i=1}^{l} L'(M_i) \frac{\partial M_i}{\partial w}$$

$$\frac{\partial M_i}{\partial w} = y_i x_i$$

$$\nabla \tilde{Q} = \sum_{i=1}^{l} y_i x_i L'(M_i)$$

$$\boxed{w_{n+1} = w_n - \gamma_n \sum_{i=1}^{l} y_i x_i L'(M_i)}$$

Стохастический градиент

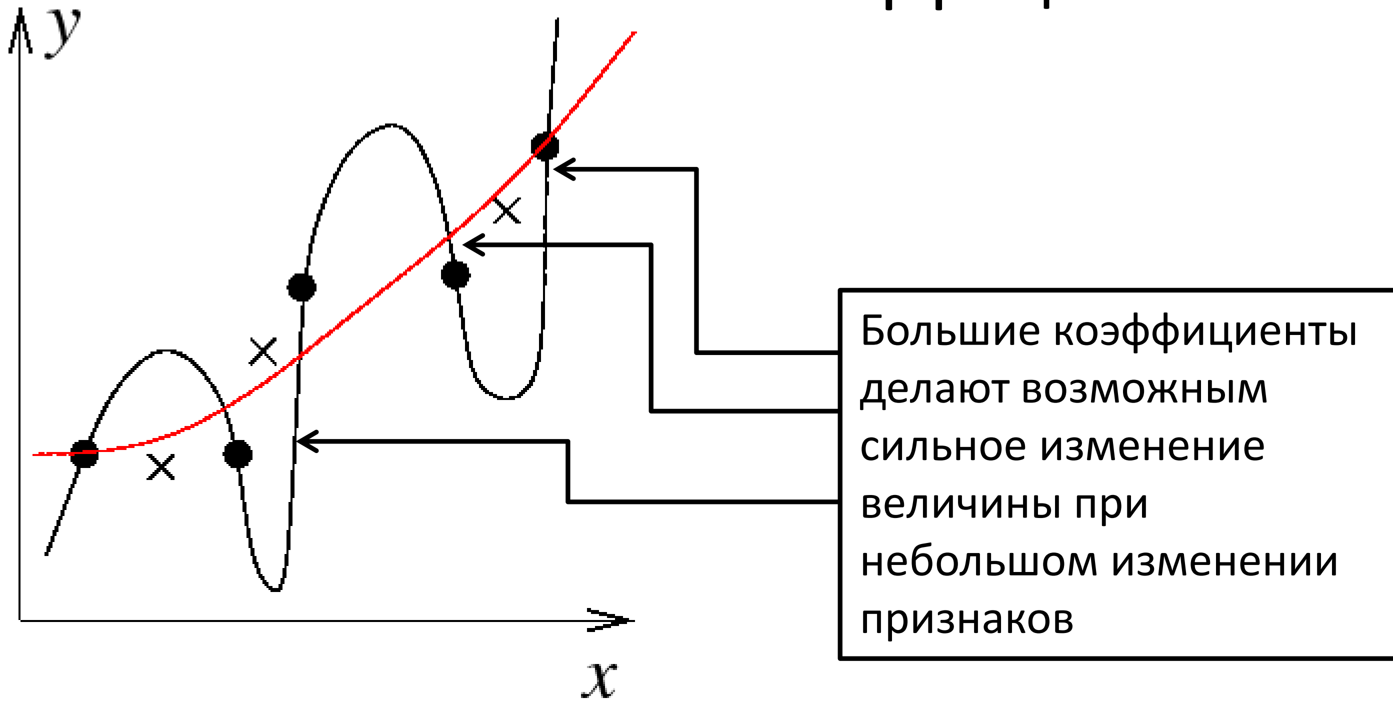
$$w_{n+1} = w_n - \gamma_n \cancel{\sum_{i=1}^{l}} y_i x_i L'(M_i)$$

$$\boxed{w_{n+1} = w_n - \gamma_n y_i x_i L'(M_i)}$$

x_i – случайный элемент обучающей выборки

Регуляризация

- Переобучение в задаче обучения с учителем как правило означает большие коэффициенты:



- Идея: добавить ограничение на коэффициенты

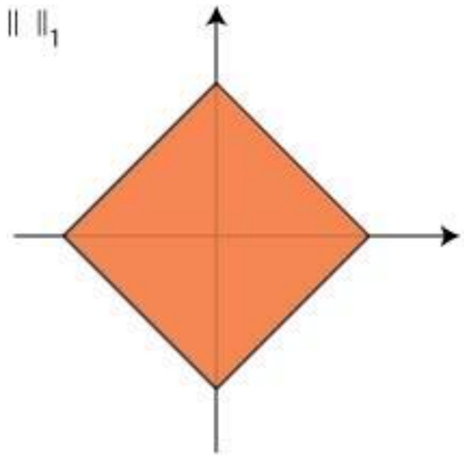
Регуляризация

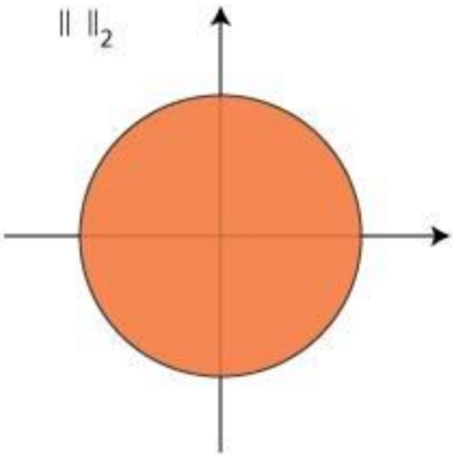
$$\begin{cases} \tilde{Q} = \sum_{i=1}^{l} L(M_i) \to min \\ \sum_{k=1}^{m} |w_k| \le \tau \end{cases}$$

$l1$ – регуляризация

$$\begin{cases} \tilde{Q} = \sum_{i=1}^{l} L(M_i) \to min \\ \sum_{k=1}^{m} {w_k}^2 \le \tau \end{cases}$$

$l2$ – регуляризация



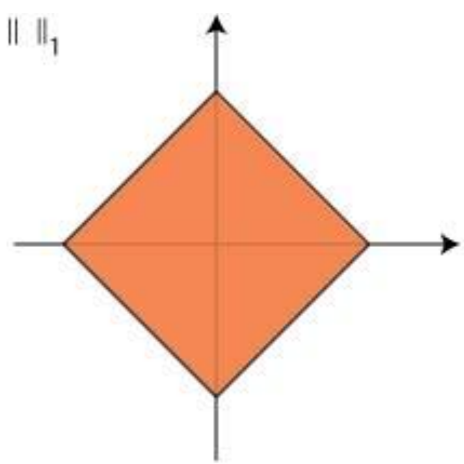


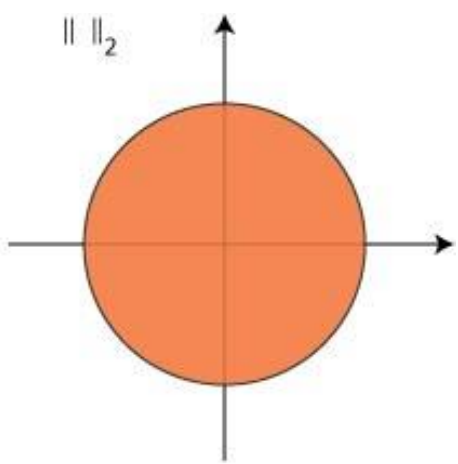
Регуляризация

$$\sum_{i=1}^{l} L(M_i) + \gamma \sum_{k=1}^{m} |w_k| \to min \qquad \sum_{i=1}^{l} L(M_i) + \gamma \sum_{k=1}^{m} {w_k}^2 \to min$$

$l1$ – регуляризация

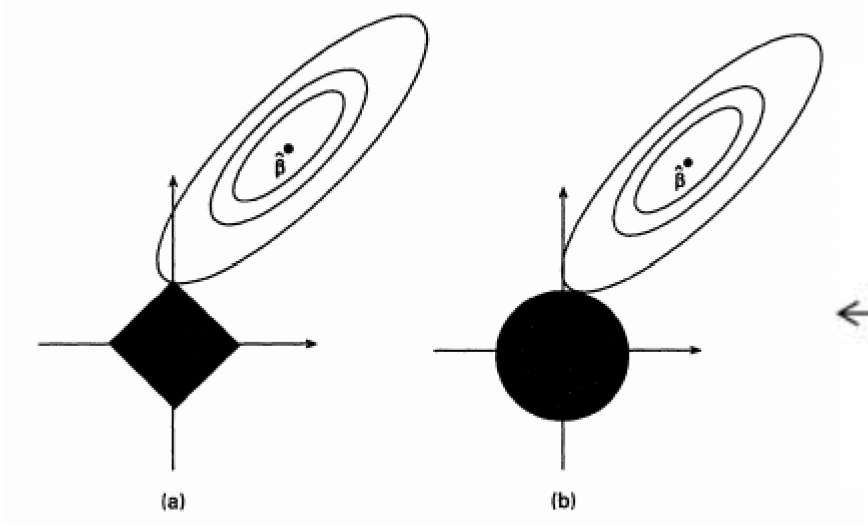
$l2$ – регуляризация

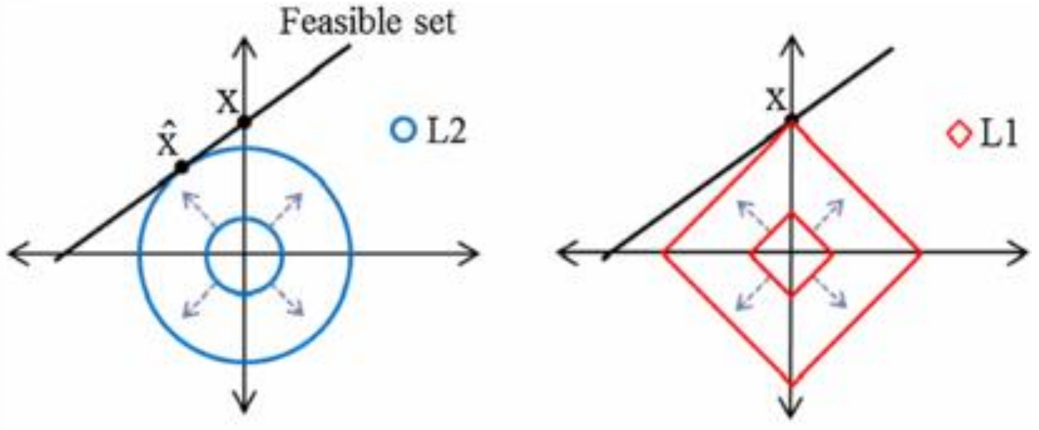




Различия между *l1* и *l2*

- Разреженность – l1-регуляризация делает вектор весов более разреженным (содержащим больше нулей)
- В случае линейной классификации это означает отбор признаков: признаки с нулевыми весами не используются в классификации
- Простая геометрическая иллюстрация разреженности:





Стандартные линейные классификаторы

Классификатор	Функция потерь	Регуляризатор
SVM (Support vector machine, метод опорных векторов)	$L(M) = \max\{0, 1 - M\}$	$\sum_{k=1}^{m} {w_k}^2$
Логистическая регрессия	$L(M) = \log(1 + e^{-M})$	Обычно $\sum_{k=1}^{m} {w_k}^2$ или $\sum_{k=1}^{m} \lvert w_k \rvert$
LDA (Линейный дискриминантный анализ/линейный дискриминант Фишера)	$L(y, f(x)) = (y - f(x))^2$ $y \in \left\{-\frac{l}{l_-}, +\frac{l}{l_-}\right\}$	$\sum_{k=1}^{m} {w_k}^2$ в случае плохой обусловленности

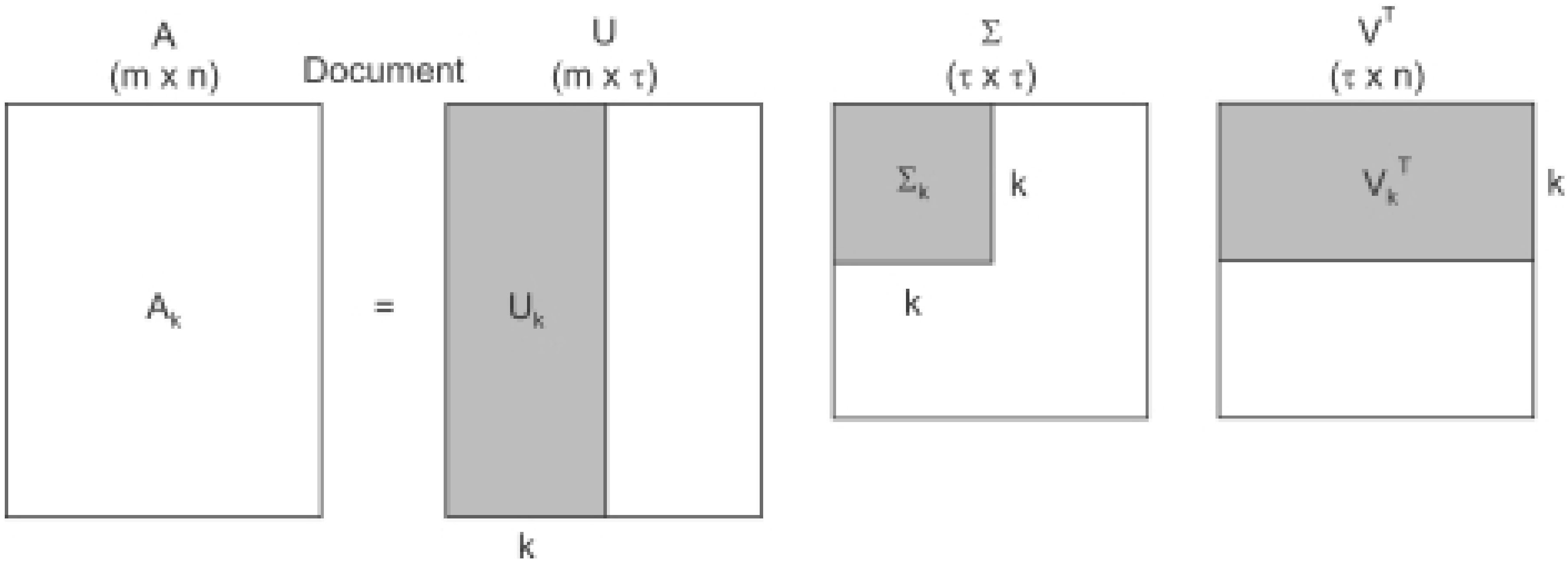
Бонусные слайды

SVD = Singular Vector Decomposition (сингулярное разложение матриц)

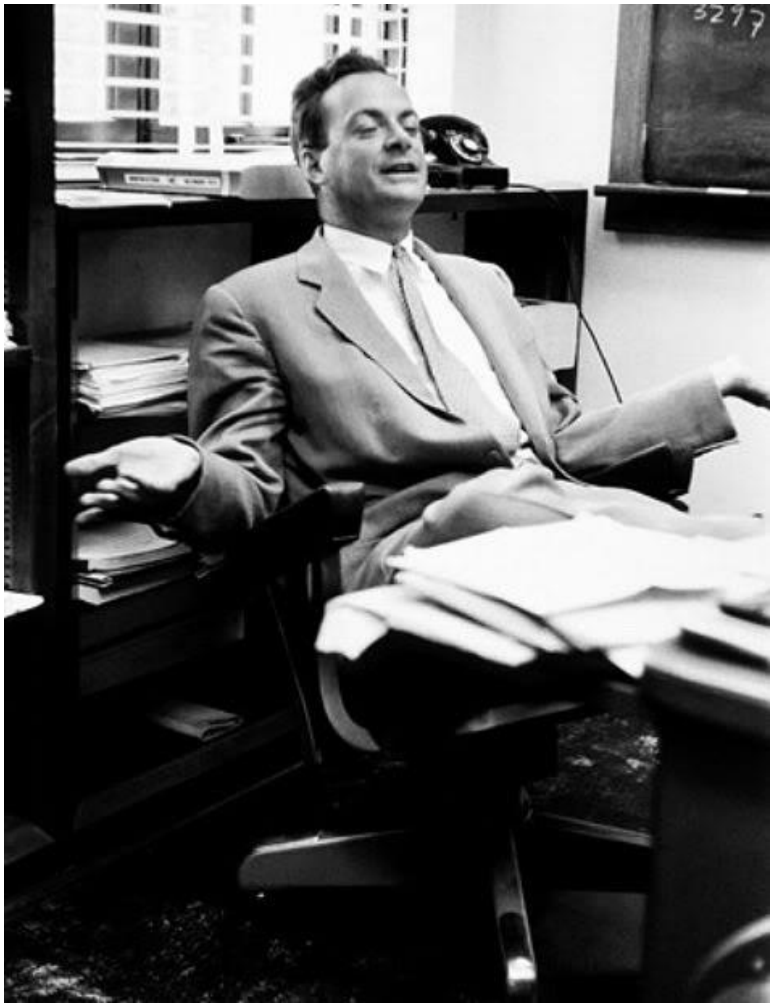
Позволяет получить наилучшее приближение исходной матрицы X матрицей X’ ранга k.

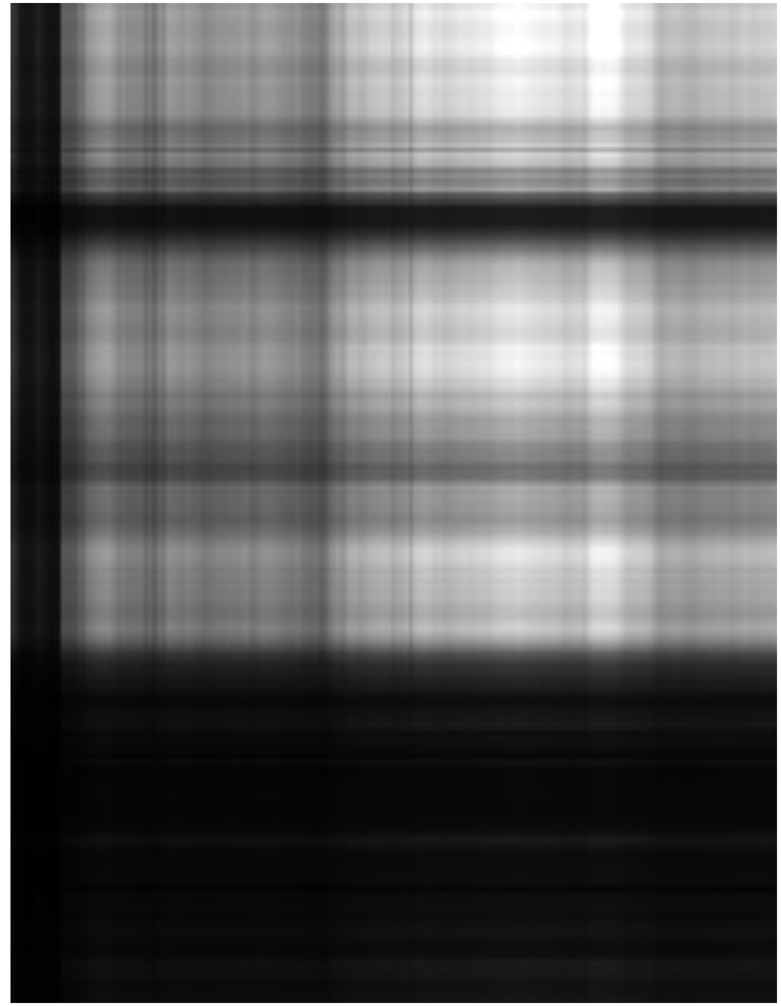
Применяется для снижения размерности пространства признаков.

SVD

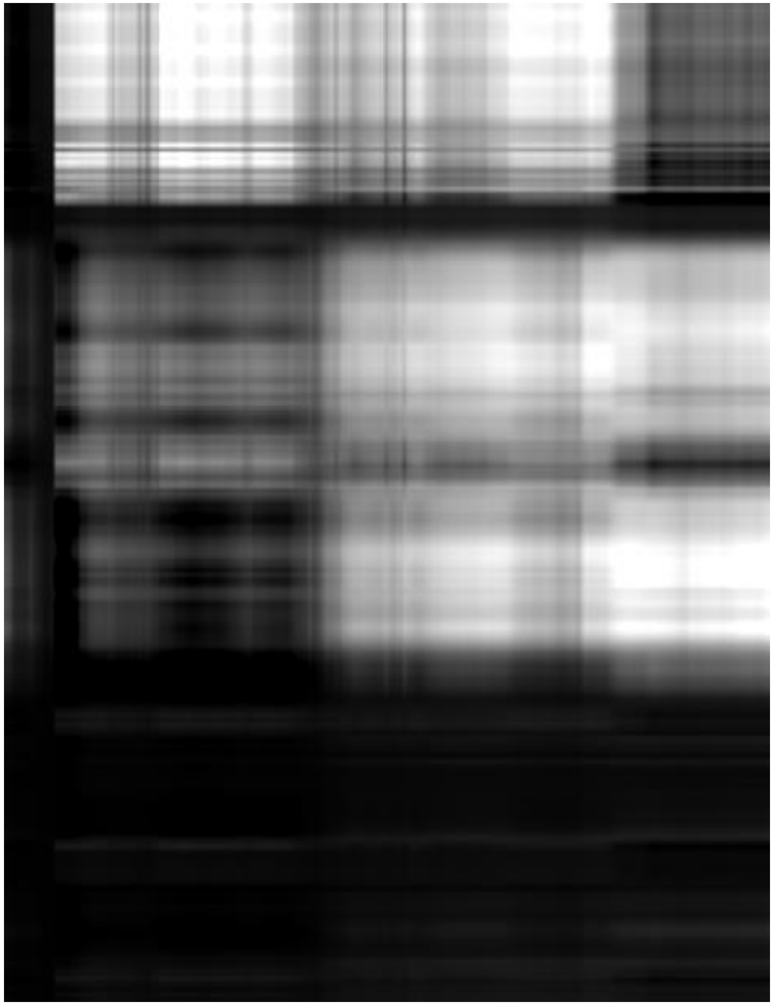


SVD: пример



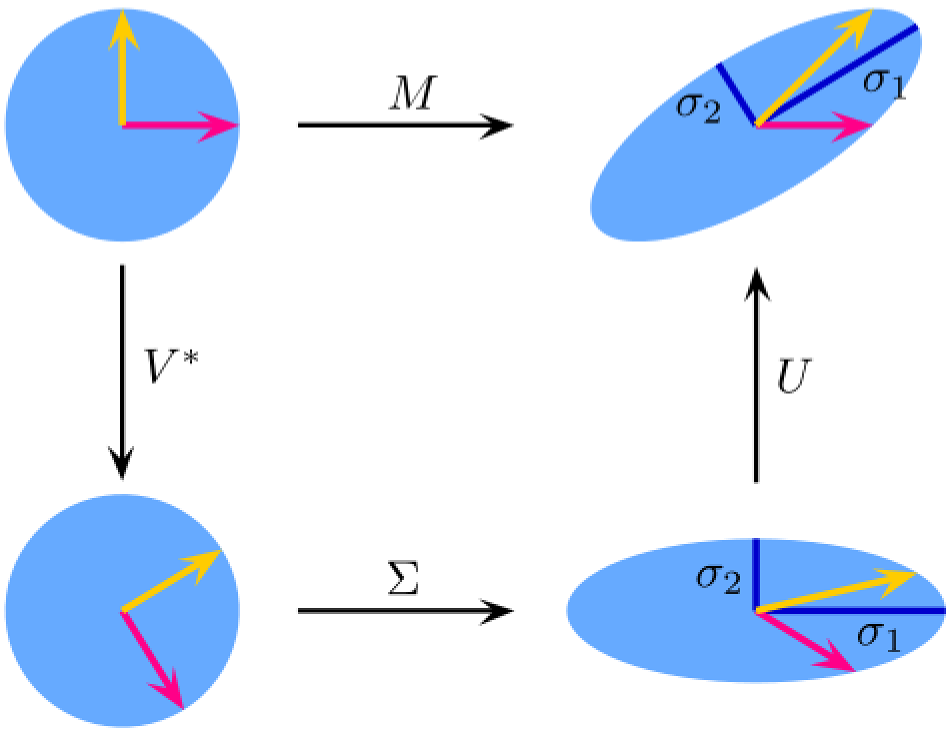


SVD: пример



SVD: пример

Геометрический смысл SVD



$$M = U \cdot \Sigma \cdot V^*$$

Итог

- Идея линейной классификации
- Функции потерь
- Градиентный спуск и стохастический градиент
- Регуляризация
- Стандартные линейные классификаторы