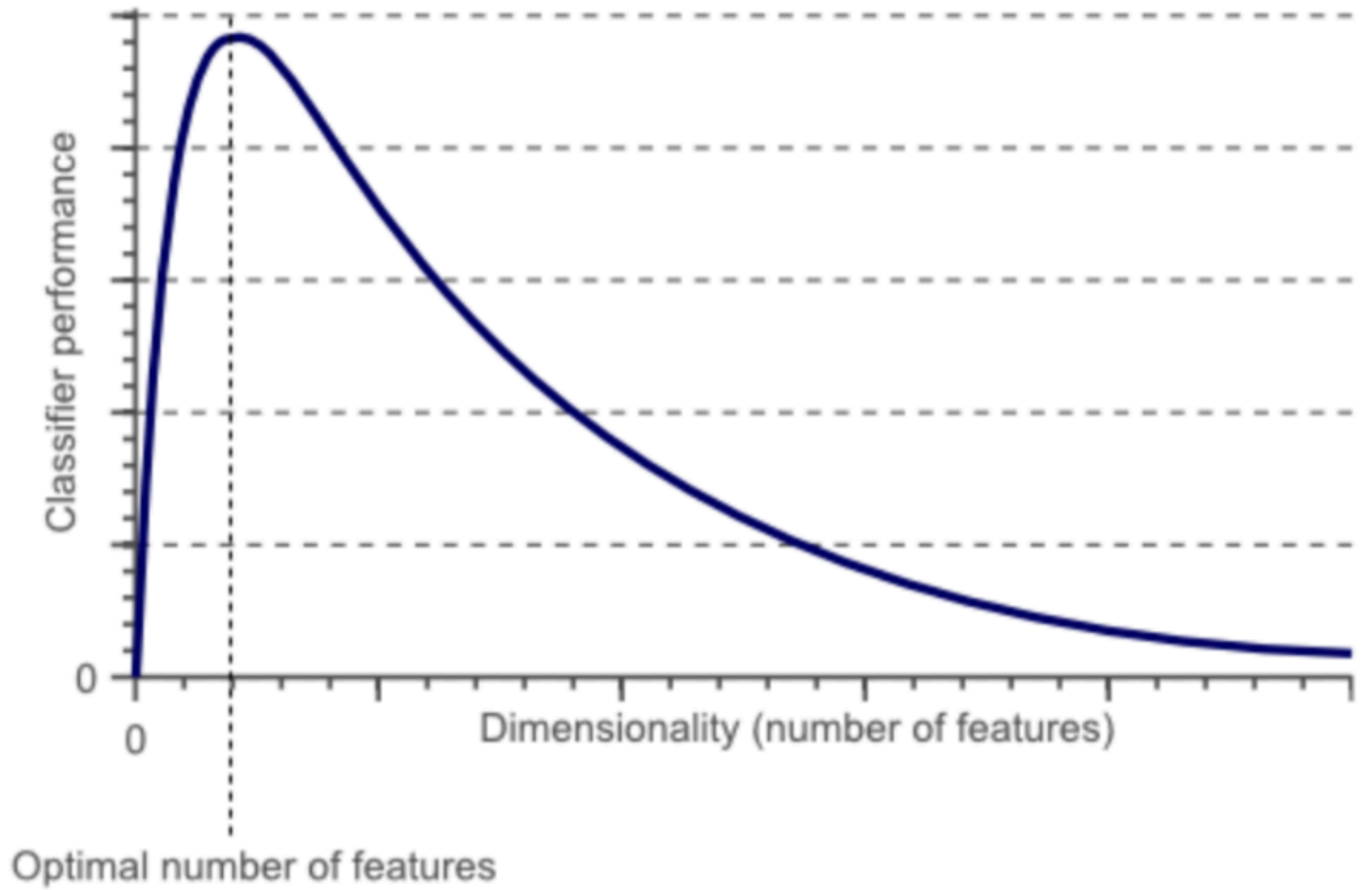
Занятие № 11

Feature Selection

Сапрыкин Артур
Data Scientist

Содержание

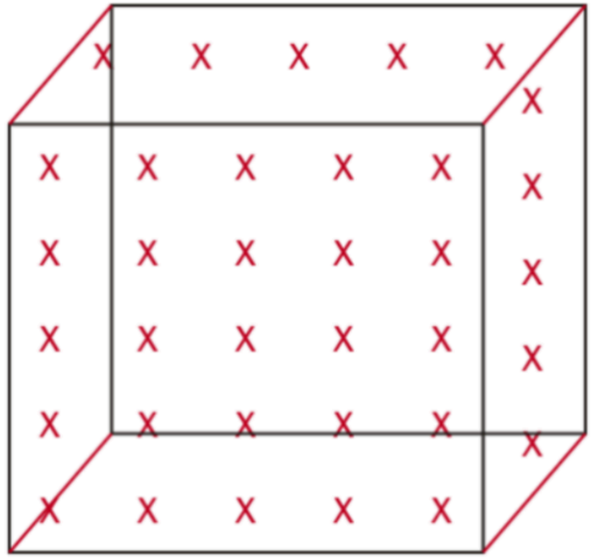
Введение. Зачем всё это?



Зачем всё это? Проклятье размерности

Одно измерение - 5 точек

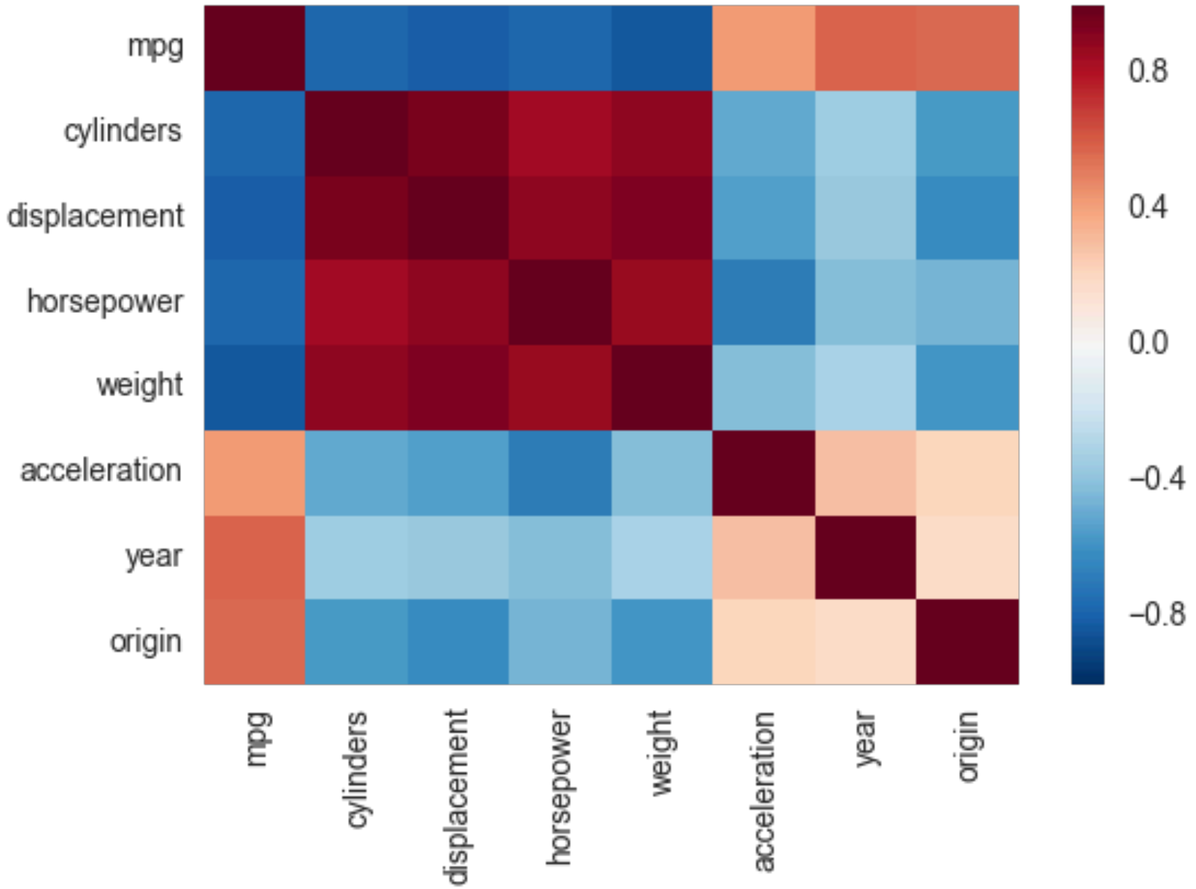
Два измерение - 25 точек



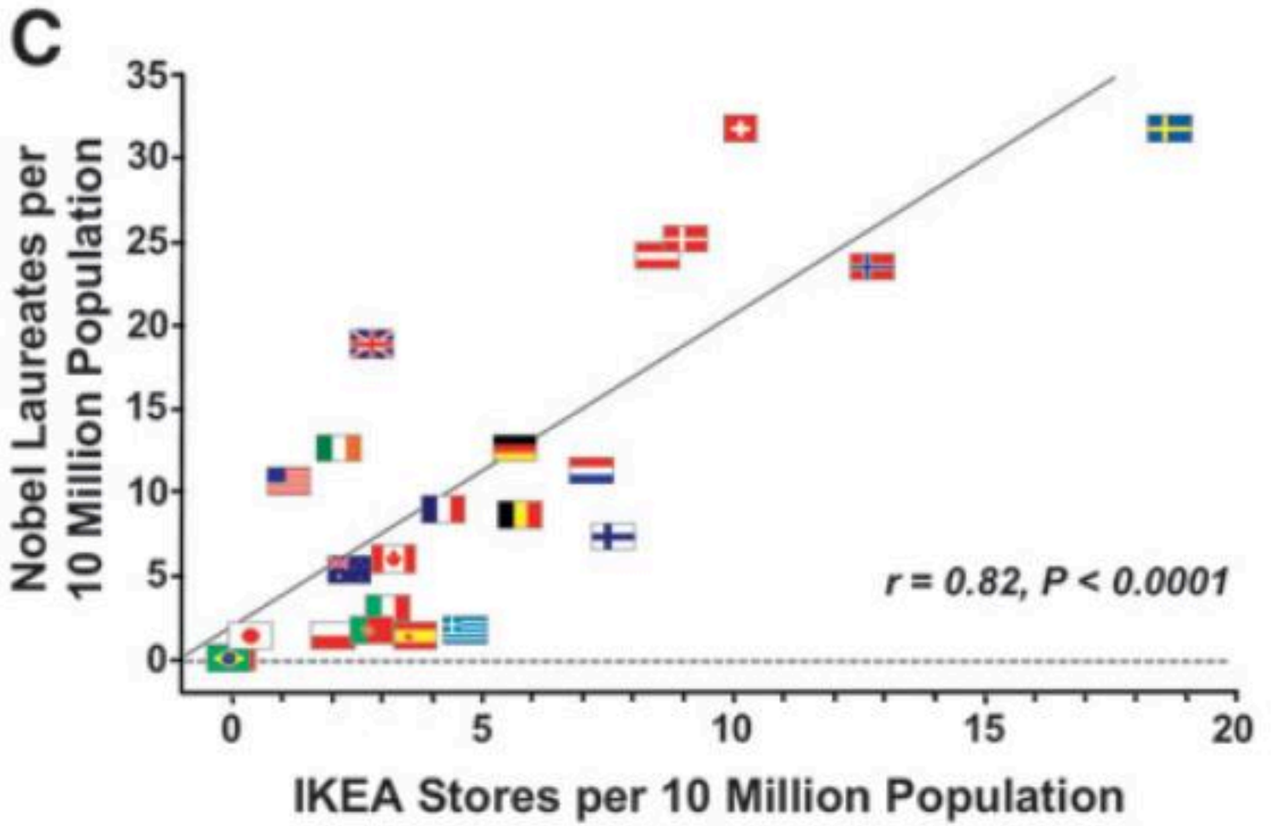
Три измерения - 125 точек

Статистика в отборе признаков

Корреляция


mpg
cylinders
displacement
horsepower
weight
acceleration
year
origin
0.8
0.4
0.0
−0.4
−0.8

Корреляция

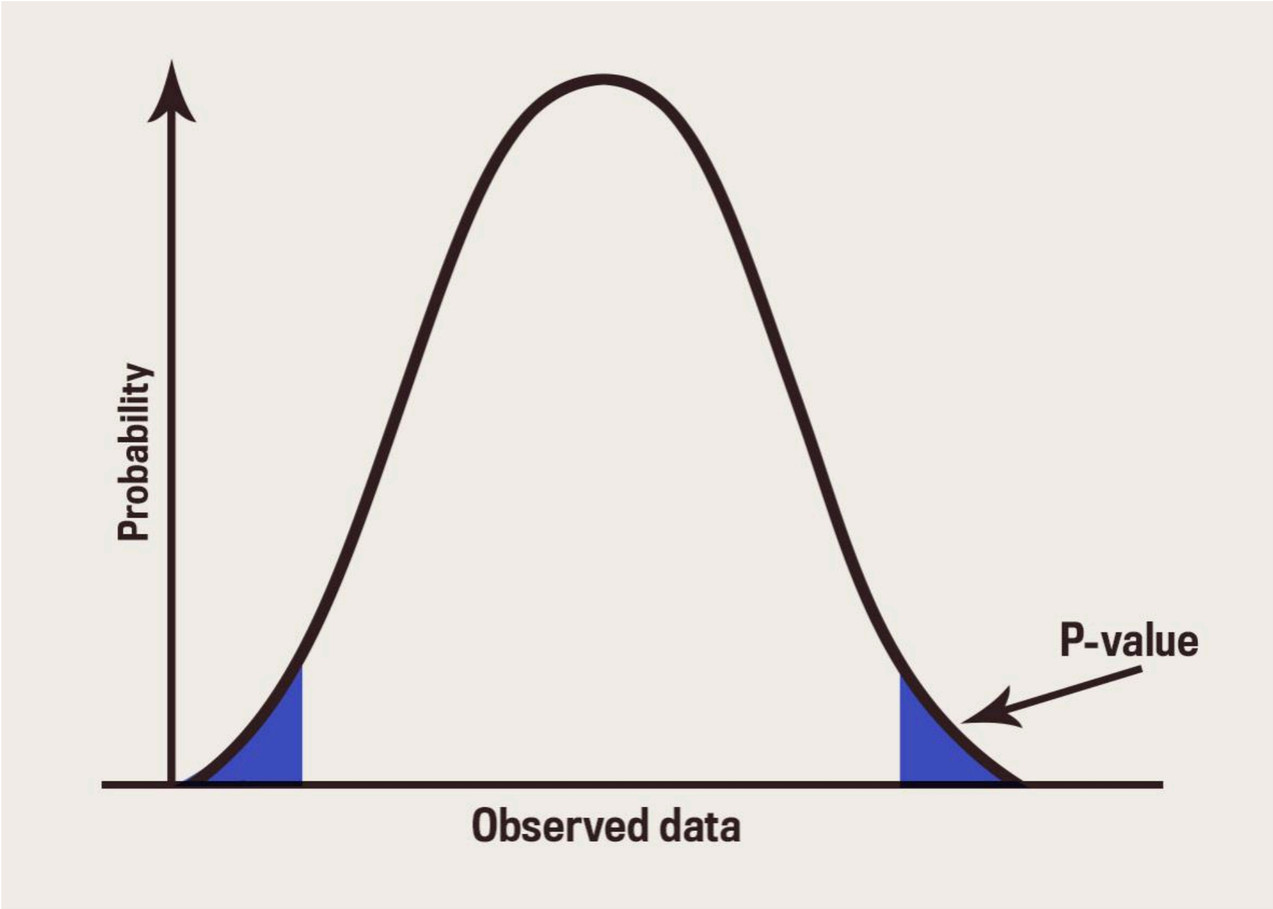

C
Nobel Laureates per 10 Million Population
35
30
25
20
15
10
5
0
0
5
10
15
20
IKEA Stores per 10 Million Population
r = 0.82, P < 0.0001
ForexAW.com

T-статистика

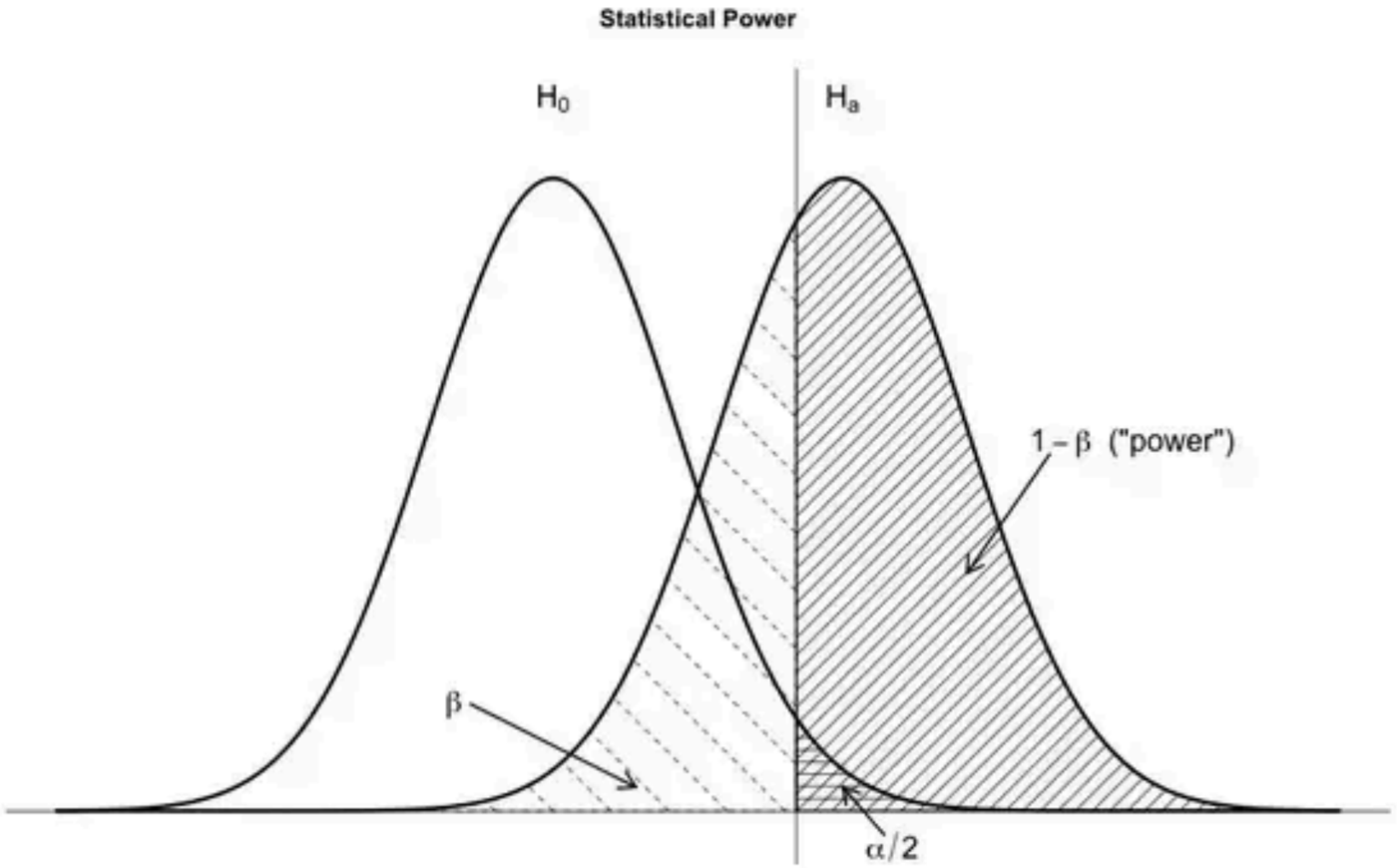
$$t = \frac{\hat{\beta}_i - 0}{SE(\hat{\beta}_i)}$$

- Если между x_i и y нет зависимости, то t соответсвует t-распределению с n−2 степенями свободы
- p-value - вероятность того, что при известном распределении наблюдаемое значение ≥ |t| (**при условии, что $\beta_i = 0$**)
- Если p-value достаточно маленький (< 1%), то мы можем отклонить H_0

P-value



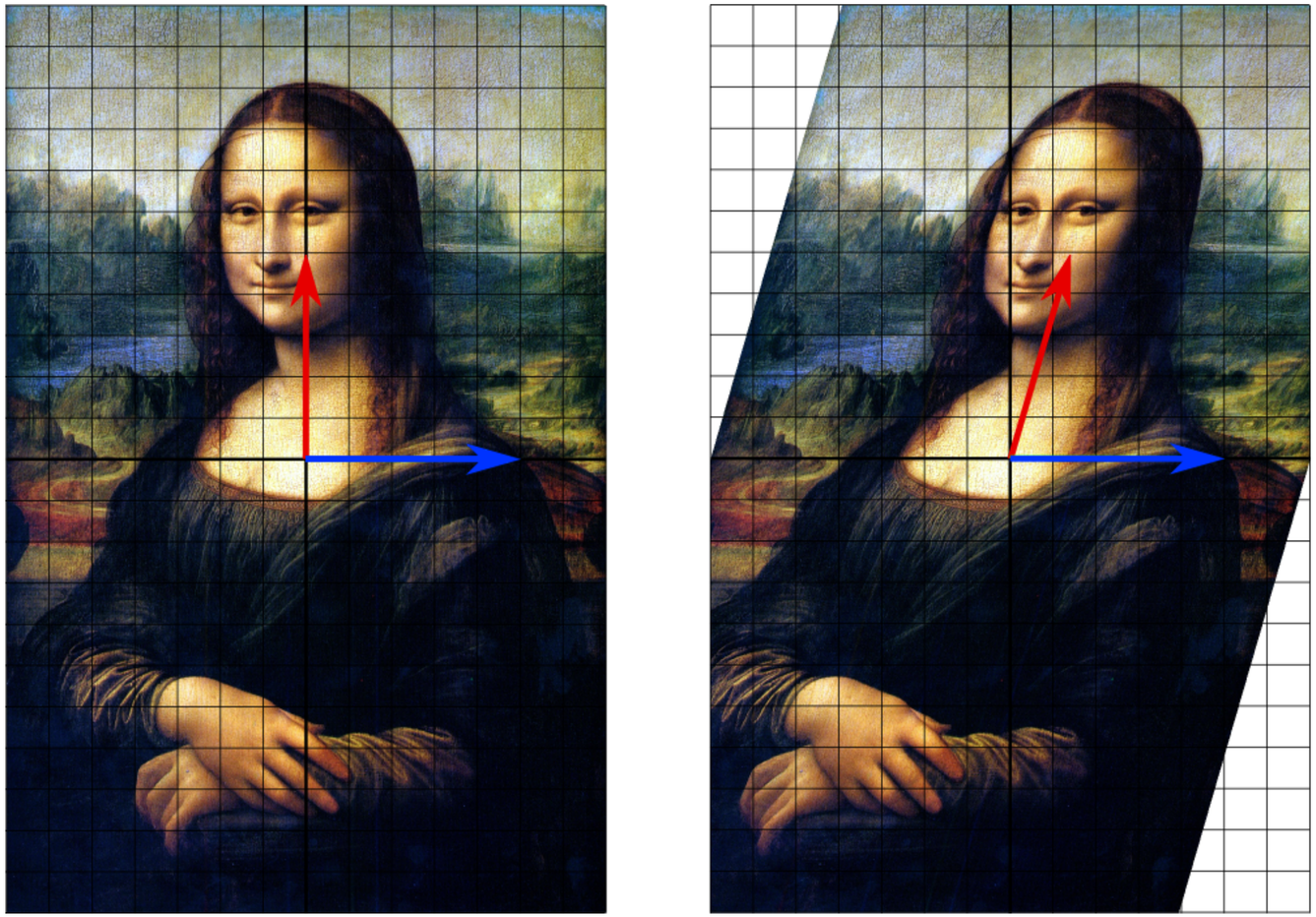
P-value



Декомпозиция данных

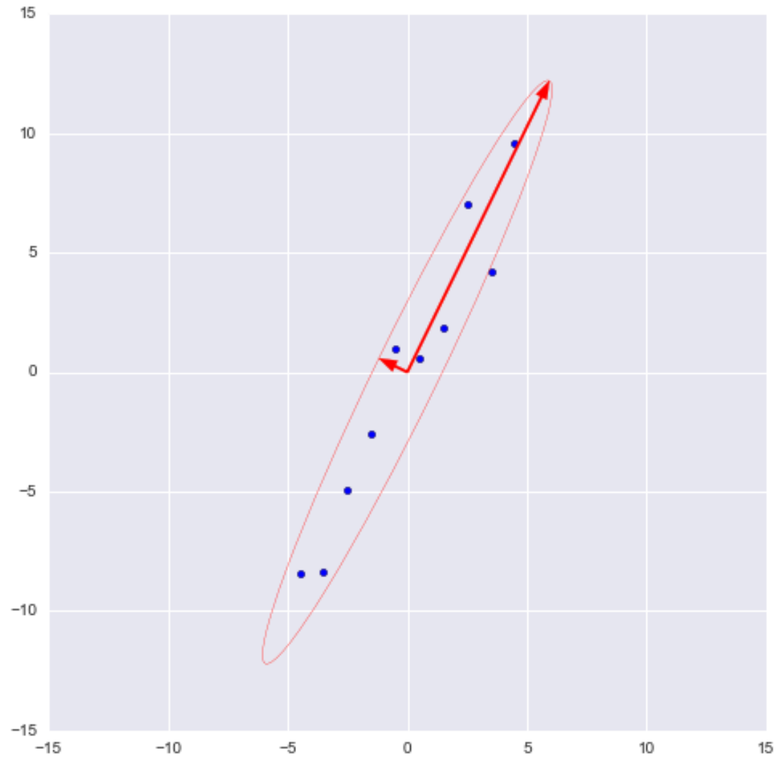
Собственный вектор

$$M\vec{x} = \lambda\vec{x}$$



PCA

Зачем он нужен? Он уменьшает размерность! ☺



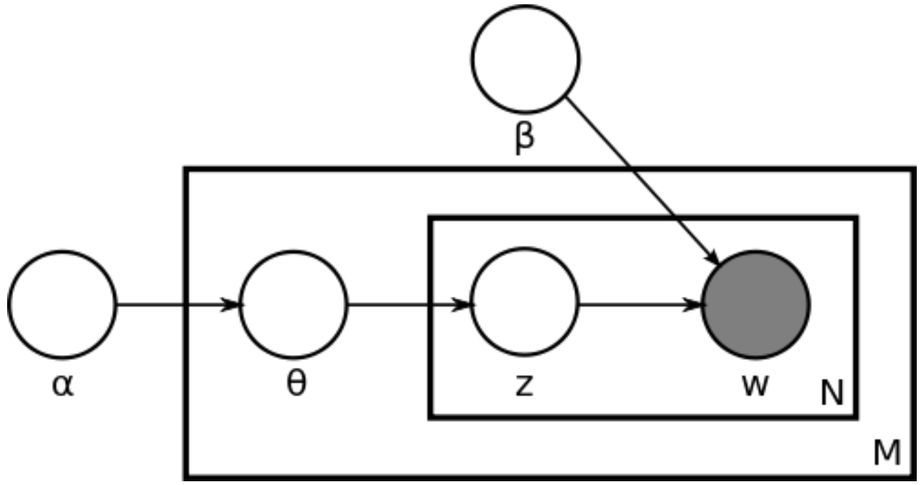
PCA

$$Cov(X_i, X_j) = E\Big[\big(X_i - E(X_i)\big) \cdot \big(X_j - E(X_j)\big)\Big] = E(X_i X_j) - E(X_i) \cdot E(X_j)$$

$$\begin{aligned} Var(X^*) = \Sigma^* &= E(X^* \cdot X^{*T}) = E\Big((\vec{v}^T X) \cdot (\vec{v}^T X)^T\Big) = \\ &= E(\vec{v}^T X \cdot X^T \vec{v}) = \vec{v}^T E(X \cdot X^T)\vec{v} = \vec{v}^T \Sigma \vec{v} \end{aligned}$$

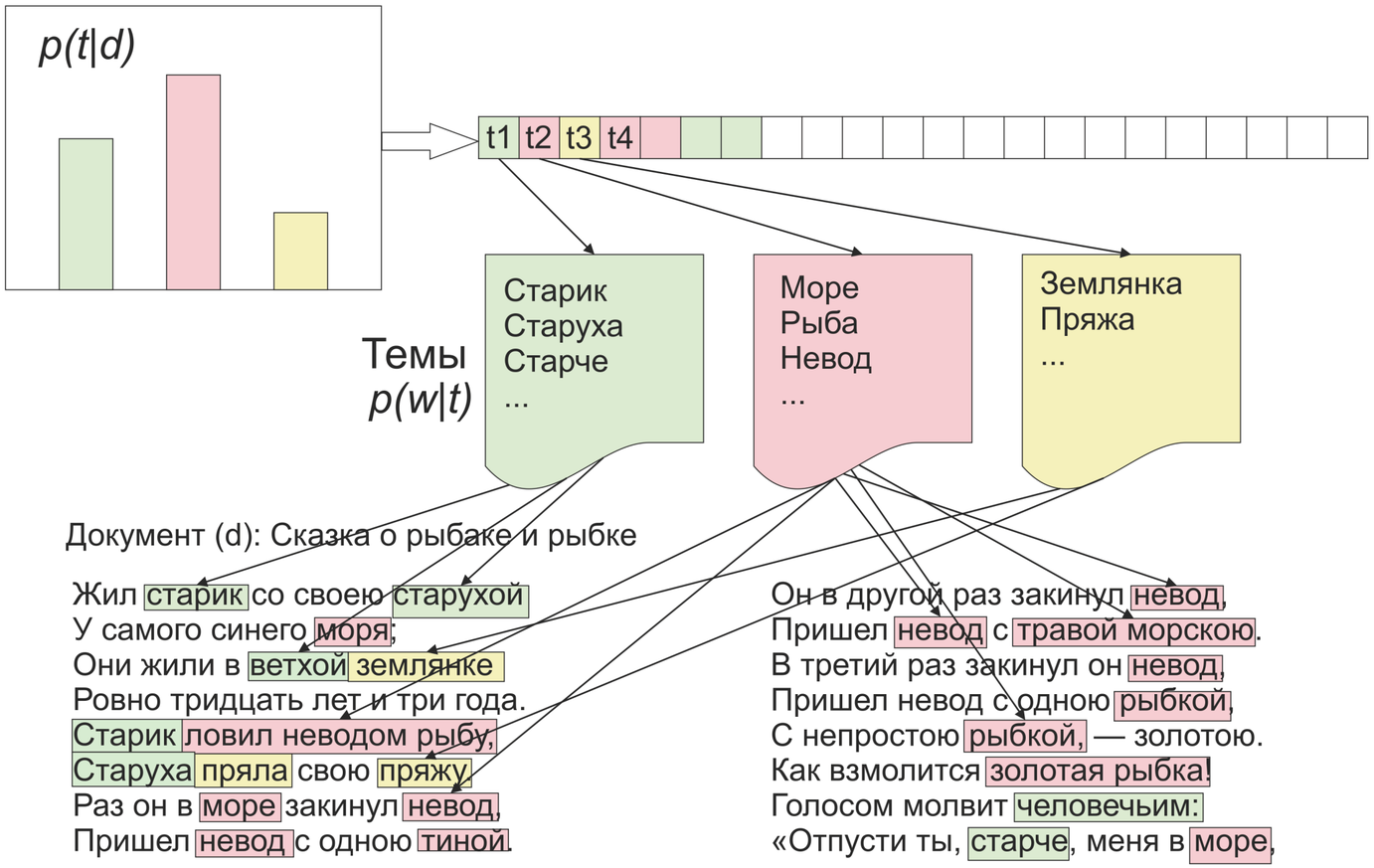
LDA

это иерархическая байесовская модель, состоящая из двух уровней:
на первом уровне – смесь, компоненты которой соответствуют «темам»;
на втором уровне – мультиномиальная переменная с априорным распределением Дирихле, которое задаёт «распределение тем» в документе.



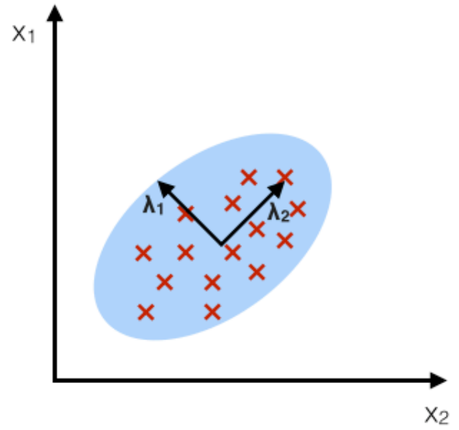
$$p(\theta, , , N \mid \alpha, \beta) = p(N \mid \xi) p(\theta \mid \alpha) \prod_{n=1}^{N} p(z_n \mid \theta) p(w_n \mid z_n, \beta).$$

LDA

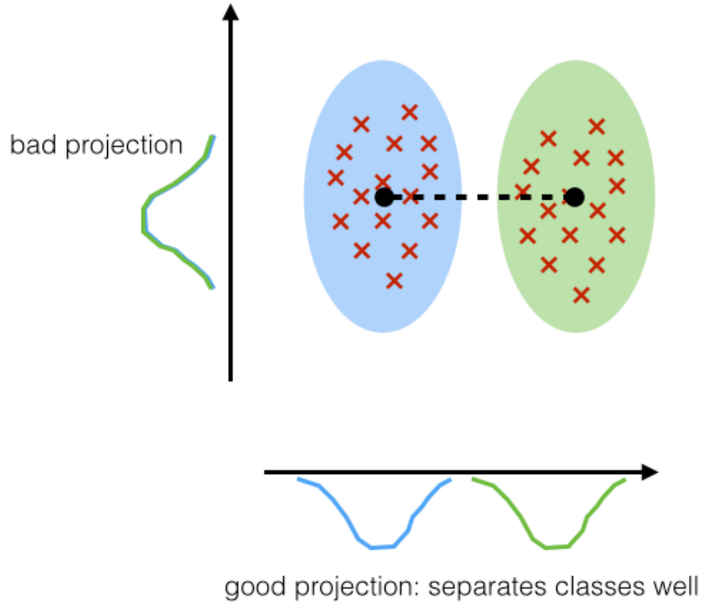

p(t|d)
t1 t2 t3 t4
Темы
p(w|t)
Старик
Старуха
Старче
...
Море
Рыба
Невод
...
Землянка
Пряжа
...
Документ (d): Сказка о рыбаке и рыбке
Жил старик со своею старухой
У самого синего моря;
Они жили в ветхой землянке
Ровно тридцать лет и три года.
Старик ловил неводом рыбу,
Старуха пряла свою пряжу.
Раз он в море закинул невод,
Пришел невод с одною тиной.
Он в другой раз закинул невод,
Пришел невод с травой морскою.
В третий раз закинул он невод,
Пришел невод с одною рыбкой,
С непростою рыбкой, — золотою.
Как взмолится золотая рыбка!
Голосом молвит человечьим:
«Отпусти ты, старче, меня в море,

Сравнение LDA и PCA

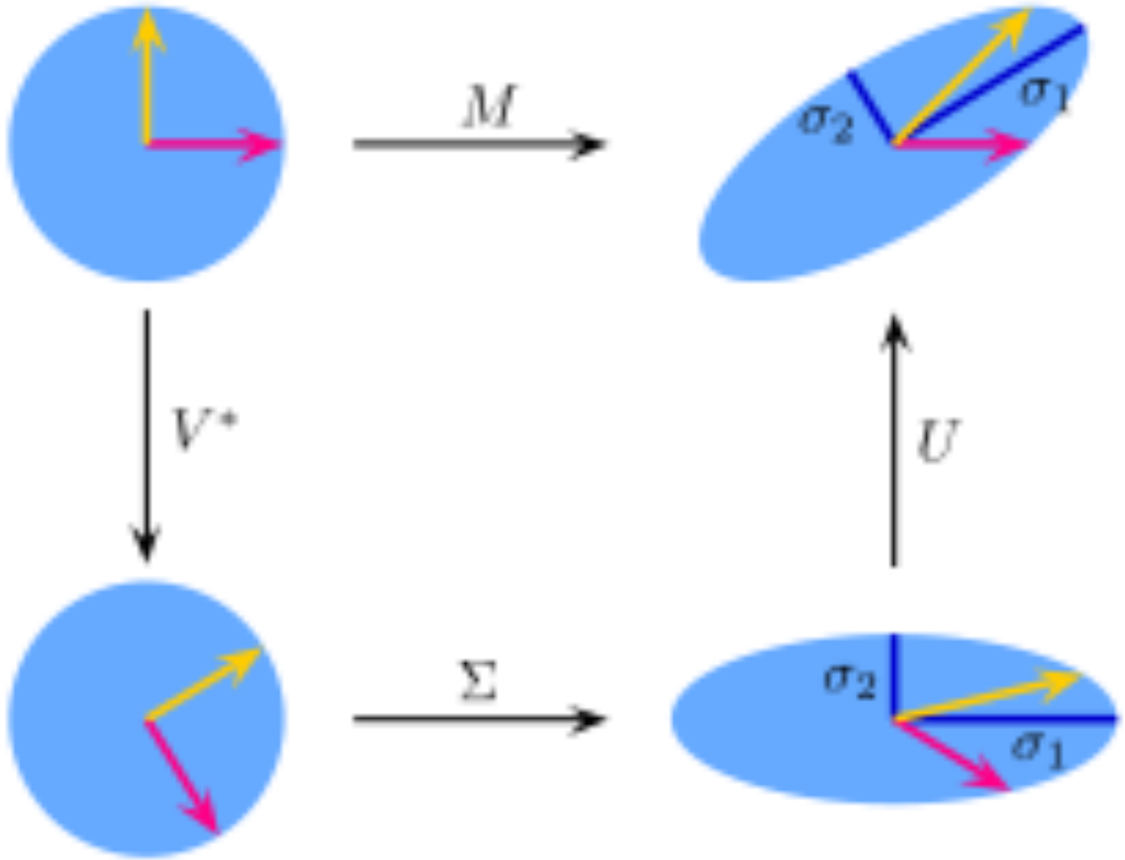
PCA:
component axes that maximize the variance



LDA:
maximizing the component axes for class-separation

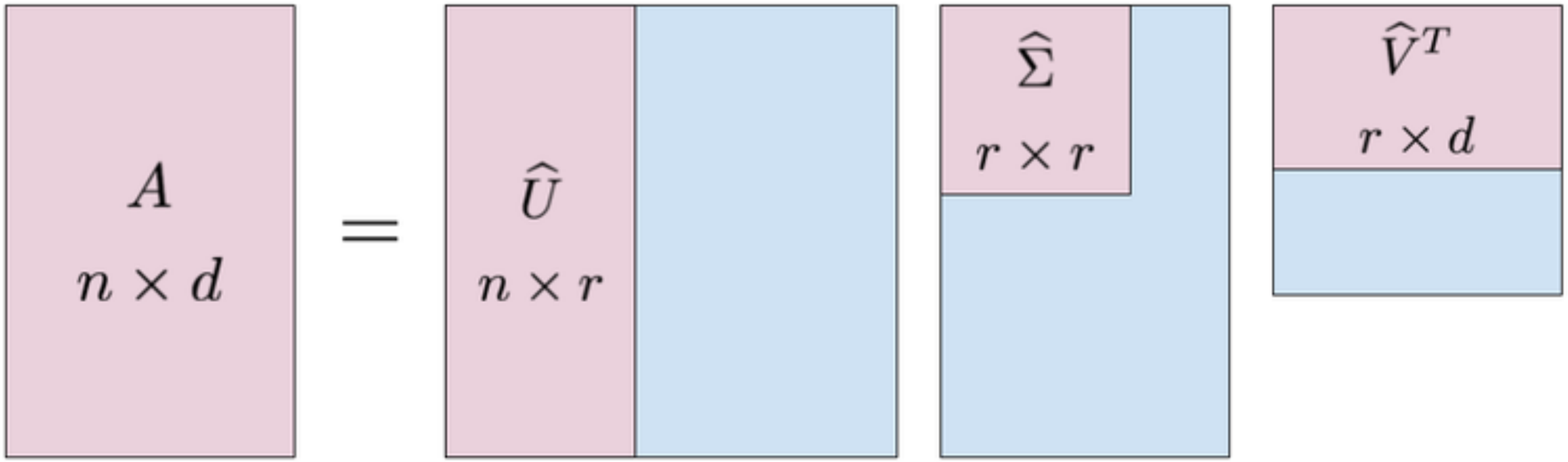


SVD



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

SVD


A
$n \times d$
$=$
$\widehat{U}$
$n \times r$
$\widehat{\Sigma}$
$r \times r$
$\widehat{V}^T$
$r \times d$
U
$n \times n$
Σ
$n \times d$
V^T
$d \times d$

ПРАКТИКА

Спасибо за внимание!

Сапрыкин Артур
Data Scientist

fb.com/asaprykin92

asaprykin92@gmail.com