Разбор домашнего задания №4 Тренировки по ML

Young&&Yandex ШАД

girafe ai

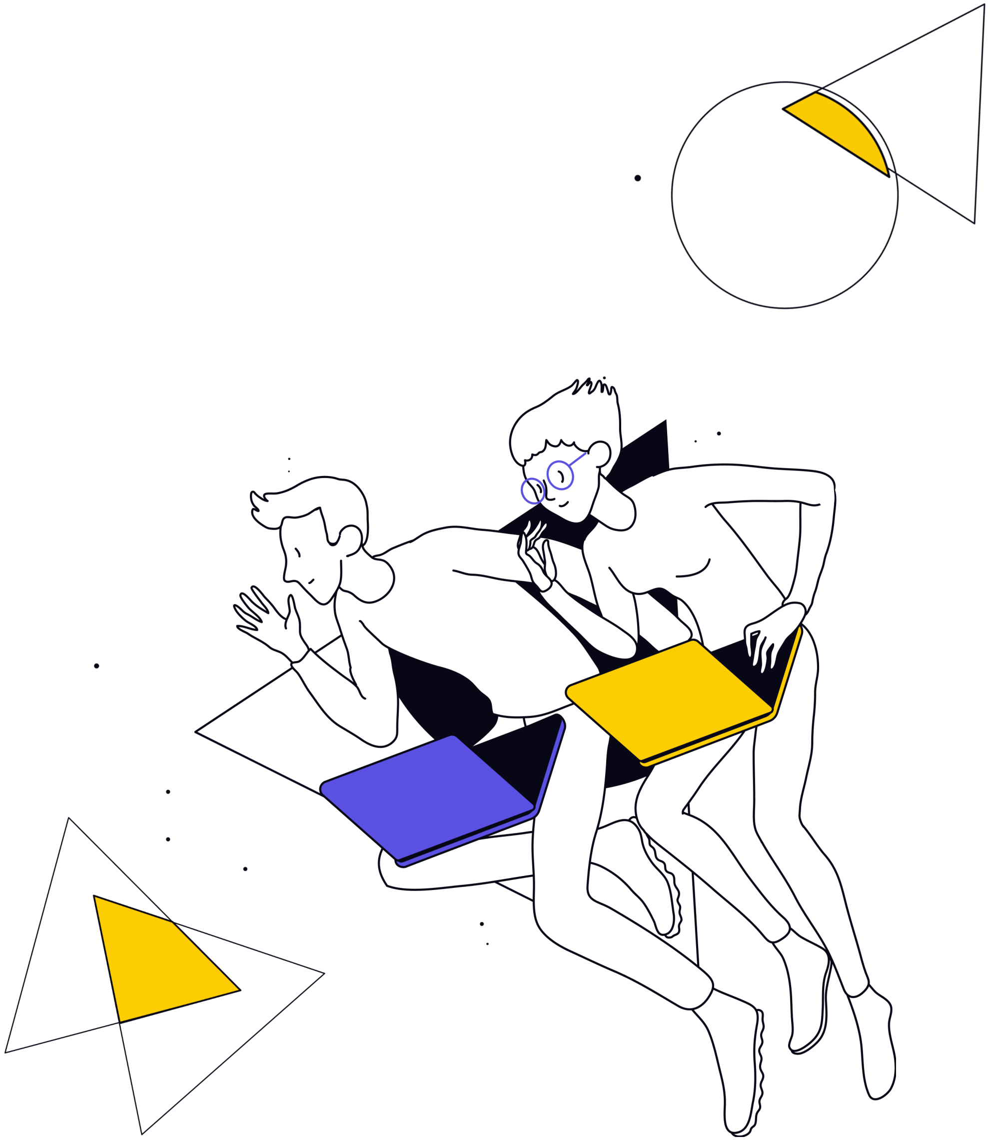
Arkadii Lysiakov

Outline

01 Reminder of models

02 Feature selection

03 Baseline review



Linear regression

Linear regression problem statement:

01 Dataset $\{(\boldsymbol{x}^{(i)}, y^{(i)})\}_{i=1}^{n}$, where $\boldsymbol{x}^{(i)} \in \mathbb{R}^p, y^{(i)} \in \mathbb{R}$

02 The model is linear:

$$\hat{y} = w_0 + \sum_{k=1}^{p} x_k w_k = //\boldsymbol{x} = [1; x_1, \dots, x_p]// = \boldsymbol{x}^\top \boldsymbol{w}$$

, where $\boldsymbol{w} = [w_0; w_1, \dots, w_p]$ is bias term

03 Least squares method (MSE minimization) provides a solution:

$$\hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}} ||\mathbf{Y} - \mathbf{X}\boldsymbol{w}||_2^2$$

Logistic regression

$$X \in R^{n \times p}$$

$$Y \in C^n \quad \text{e.g. } C = \{-1, 1\}$$

$$|C| < +\infty$$

$$c(X) = \hat{Y} \approx Y$$

Logistic regression

Just to remind

$$\log L(w|X,Y) = \log P(X,Y|w) = \log \prod_{i=1}^{n} P(x_i, y_i|w)$$

Calculating probabilities for objects

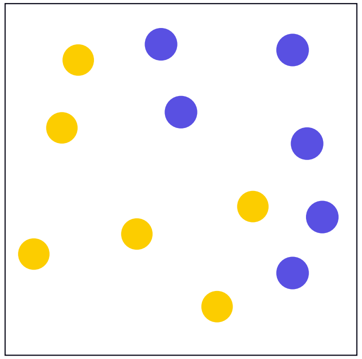
$$\text{if } y_i = 1: \quad P(x_i, 1|w) = \sigma_w(x_i) = \sigma_w(M_i)$$

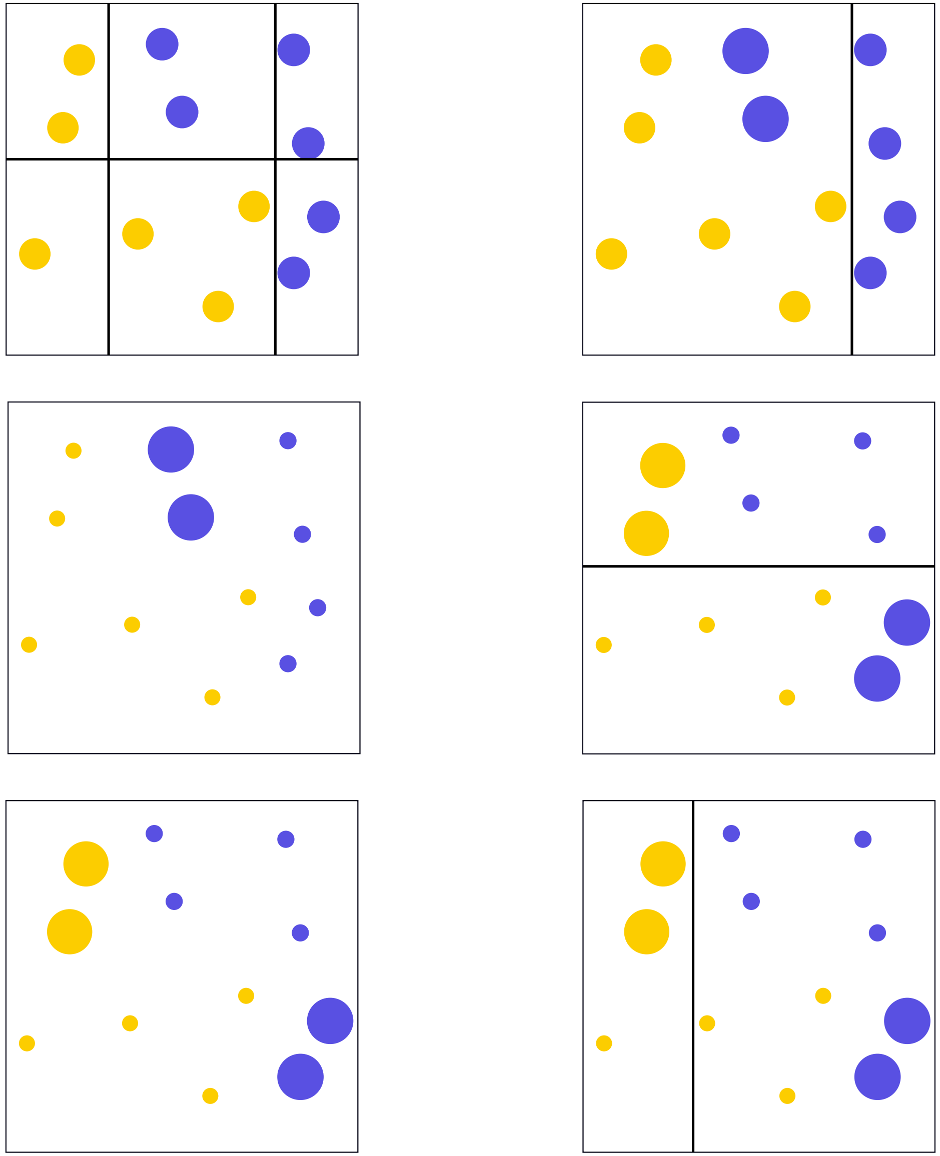
$$\text{if } y_i = -1: \quad P(x_i, -1|w) = 1 - \sigma_w(x_i) = \sigma_w(-x_i) = \sigma_w(M_i)$$

$$\log L(w|X,Y) = \sum_{i=1}^{n} \log \sigma_w(M_i) = \boxed{- \sum_{i=1}^{n} \log(1 + \exp(-M_i)) \to \max_w}$$

Boosting: intuition

Binary classification
Use decision stumps





Gradient boosting

01 $$\hat{f}_{T-1}(\boldsymbol{x}) = \sum_{t=0}^{T-1} g_t(\boldsymbol{x}),$$

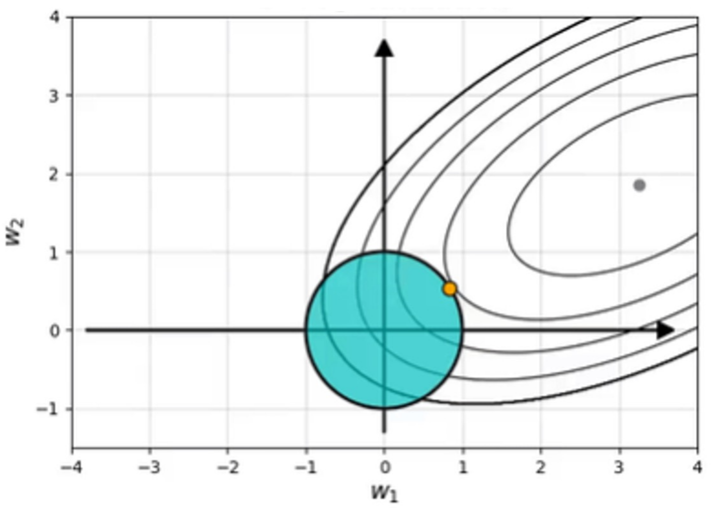
02 $$r_t^{(i)} = -\left[\frac{\partial L(y^{(i)}, f(\boldsymbol{x}^{(i)}))}{\partial f(\boldsymbol{x}^{(i)})}\right]_{f(\boldsymbol{x})=\hat{f}_T(\boldsymbol{x})}, \text{for } i = 1, \ldots, n,$$

03 $$\boldsymbol{\theta}_T = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \left(r_t^{(i)} - h(\boldsymbol{x}^{(i)}, \boldsymbol{\theta})\right)^2,$$

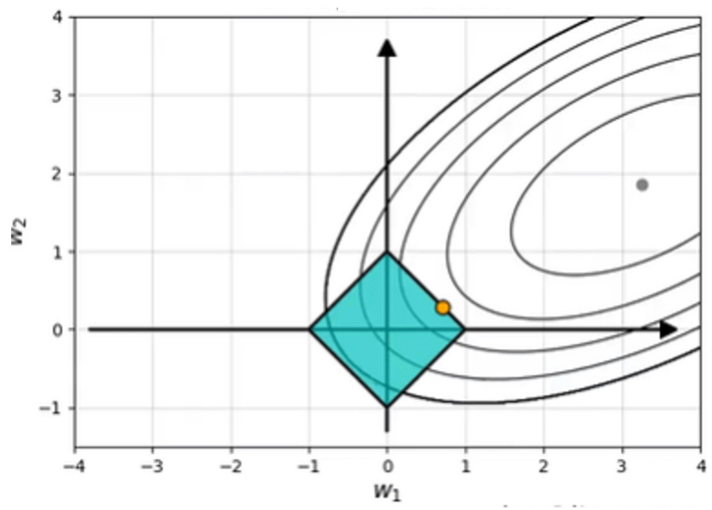
04 $$\rho_t = \arg\min_{\rho} \sum_{i=1}^{n} L\left(y^{(i)}, \hat{f}_{t-1}(\boldsymbol{x}^{(i)}) + \rho \cdot h(\boldsymbol{x}^{(i)}, \boldsymbol{\theta}_T)\right)$$

Regularization

L2 regularization $||\boldsymbol{w}||_2^2$



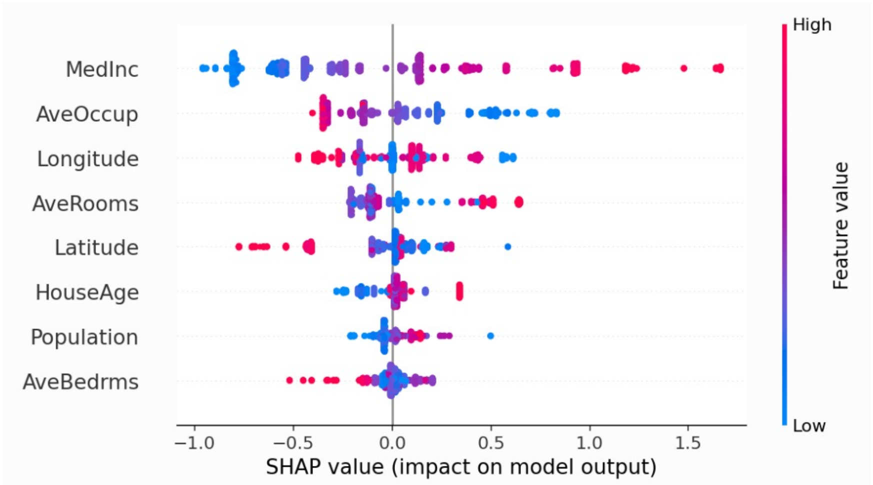
L1 regularization $||\boldsymbol{w}||_1$



source: https://people.eecs.berkeley.edu/~jrs/189/

Shap values

$$\phi_i(p) = \sum_{S \subseteq N / \{i\}} \frac{|S|!\,(n - |S| - 1)!}{n!}(p(S \cup \{i\}) - p(S))$$



$p(S \cup \{i\})$ - this is the prediction of the model with the i-th feature

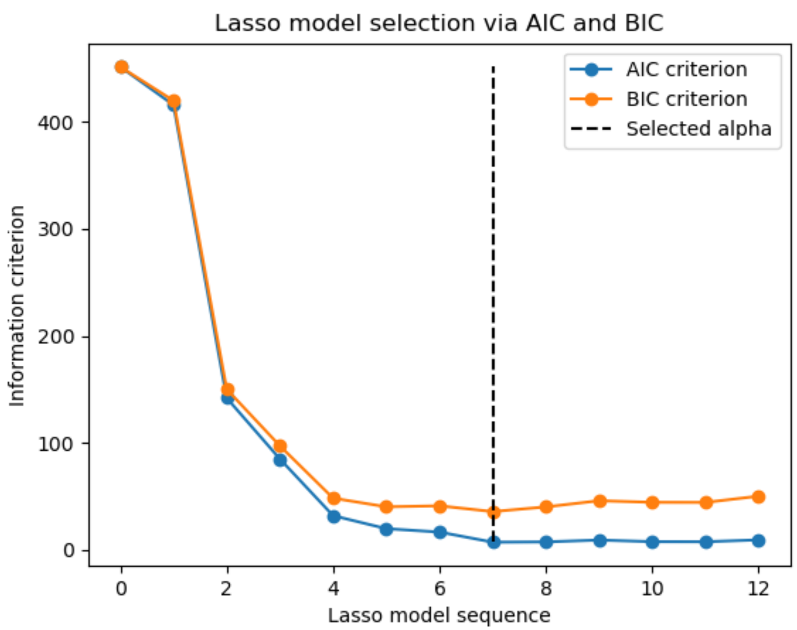
$p(S)$ - this is a prediction of the model without the i-th feature

n - number of features

S - an arbitrary set of features without the i-th feature

source: https://shap.readthedocs.io/en/latest/example_notebooks/overviews/

Information-criteria based model selection



$$AIC = -2\ \log(\hat{L}) + 2d$$

$$BIC = -2\ \log(\hat{L}) + \log(N)\ d$$

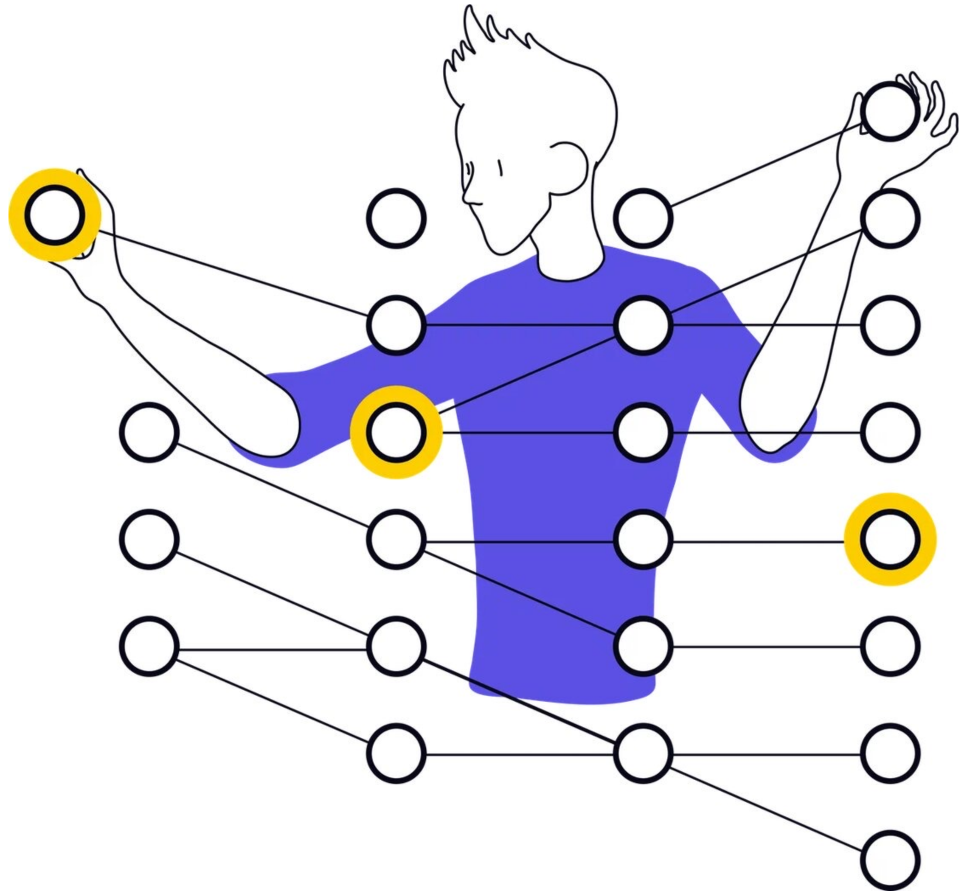
$\hat{L}$ - is the maximum likelihood of the model

d - is the number of parameters

N - is the number of samples

source: https://scikit-learn.org/dev/modules/linear_model.html

Lifecode



Thanks for attention!

Questions?

girafe ai