

Обучение с учителем: Регуляризация в линейных моделях. Метод Ближайших Соседей (KNN)

Екатерина Кондратьева

Регрессия:

МНК функция потерь:

$$\begin{aligned}\text{RSS}(\beta) &= \sum_{i=1}^N (y_i - f(x_i))^2 \\ &= \sum_{i=1}^N \left(y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2.\end{aligned}$$

N —number of samples

p —number of independent variables or features

x —feature

y —actual target or dependent variable

$f(x)$ —estimated target

β —coefficient or weight corresponding to each feature or independent var.

Регрессия

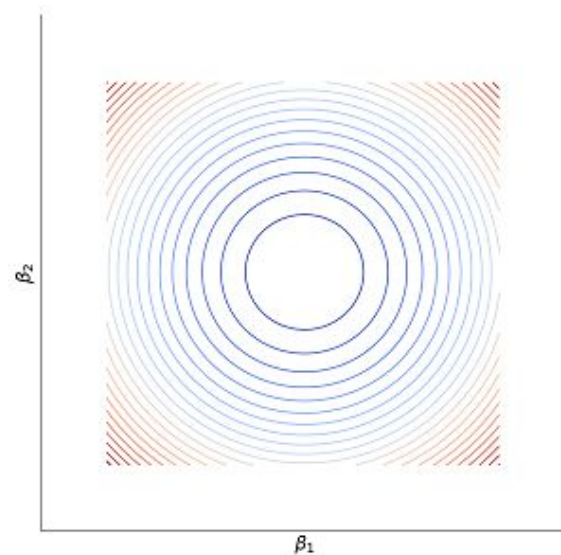
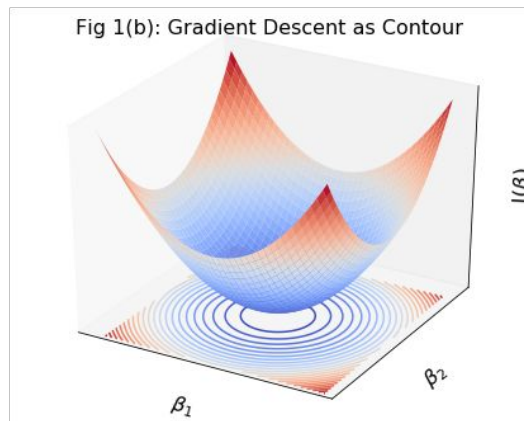
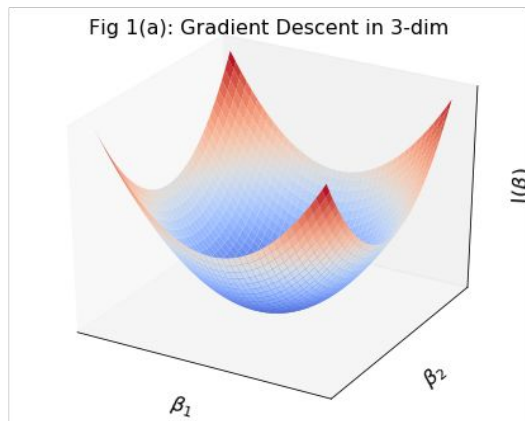


Fig 2: Gradient Descent on axes of β_1 and β_2

Регуляризация

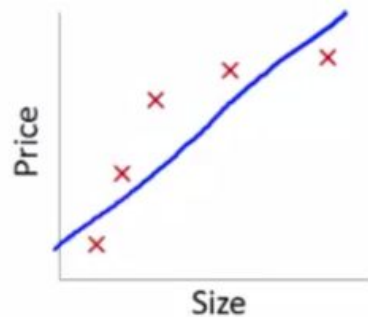
Для улучшения обобщающей способности получающейся модели, то есть уменьшения эффекта переобучения, на практике часто рассматривается логистическая регрессия с регуляризацией.

<https://docplayer.ru/41305484-Lekciya-2-obobshchennye-lineynye-modeli-regulyarizaciya-obucheniya.html>

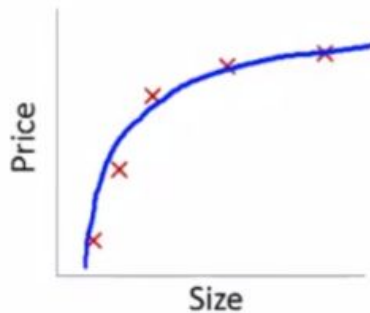
<https://github.com/esokolov/ml-course-hse/blob/master/2018-fall/lecture-notes/lecture03-linregr.pdf>

лекции https://www.youtube.com/watch?v=Kloz_aa1ed4

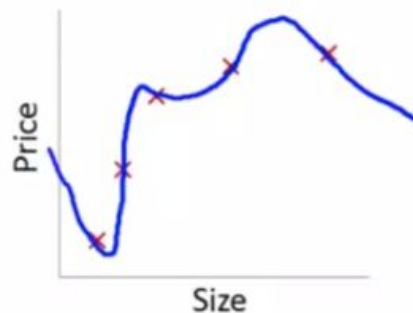
Переобучение (data leakage)



$\rightarrow \theta_0 + \theta_1 x$
"Underfit" "High bias"



$\rightarrow \theta_0 + \theta_1 x + \theta_2 x^2$
"Just right"



$\rightarrow \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$
"Overfit" "High variance"

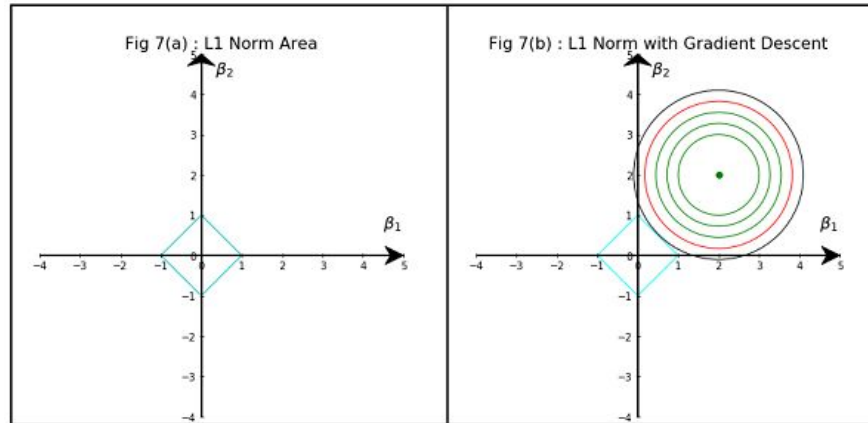
Здесь theta (θ) - β

L1 Norm or Lasso Regression

L1 Norm is of the form $|\beta_1| + |\beta_2|$.

Modified Cost function for L1 Regularization is as follows:

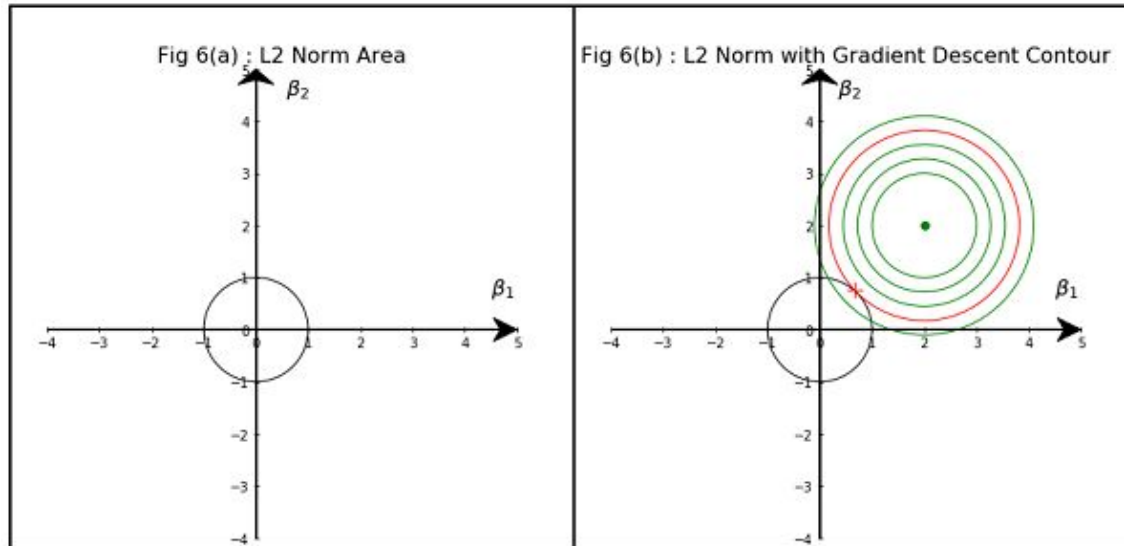
$$\hat{\beta}^{\text{lasso}} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2} \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\}$$



L2 Norm or Ridge Regression

L2 Norm is Euclidean distance norm of the form $|\beta_1|^2 + |\beta_2|^2$.

$$\hat{\beta}^{\text{ridge}} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right\}$$



Метод k Ближайших Соседей

Метод к ближайших соседей



Метод k ближайших соседей

KNN NEIGHBORHOOD
SIZE

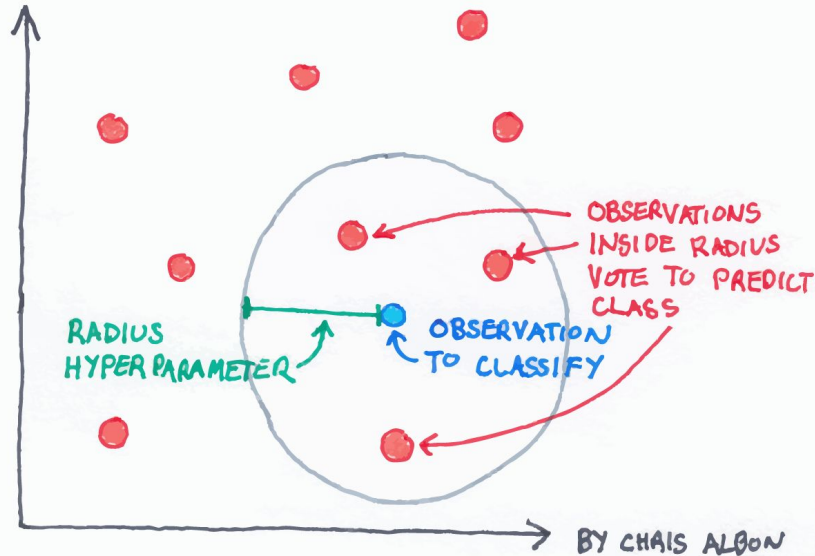
Small
↓
 $K = \text{Low Bias, High Variance}$

↑
LARGE
 $K = \text{High Bias, Low Variance}$

BY CHRIS ALBON

RADIUS-BASED NEAREST NEIGHBOR CLASSIFIER

An alternative to
k-nearest neighbor
wherein the nearest
neighbor is determined
by a radius hyper-
parameter.



Часто метрики дистанции используются для снижения размерности:

<https://www.stat.berkeley.edu/~bickel/mldim.pdf>

KNN BE LIKE



**"Show me your
friends, and I'll tell
you who you are."**

Вопросы для самопроверки:

- Почему L1-регуляризация производит отбор признаков?
- Почему коэффициент регуляризации нельзя подбирать по обучающей выборке?
- Почему категориальные признаки нельзя закодировать натуральными числами? Что такое one-hot encoding?
- Разница MAE и MSE. Почему MSE чувствительно к выбросам?
- Что такое кросс-валидация, чем она лучше использования отложенной выборки?

Источники:

1. <https://github.com/esokolov/ml-course-hse/>
2. <https://chrisalbon.com/>
3. https://github.com/Slinkolgor/express_ml