AGH University of Science
and Technology

Seminar in *Artificial Intelligence*

Marcin Zajac, Dominik Koza, Lukasz Gorczyca

Department of Telecommunications

01.04.2019

Agenda

- 1. Intro, presentation plan.
- 2. What is regression?
- 3. What regression is used for?
- 4. Types of regression.
- 5. Linear regression.
- 6. Logistic regressiion.
- 7. Polynominal regression.
- 8. QA
- 9. Quiz.

2. What is regression?

- looks for the relationship between two or more variables.

3. What regression is used for?

- used in: forecasting, MS Excel (XD), machine learning...

4. Types of regression

- linear regression
- logistic regression
- polynominal regression
- stepwise regression
- ridge regression
- lasso regression
- elasticNet regression

5.Linear regression

- First known research in this area - method of least squares published by Legendre in 1805 and by Gauss in 1809
- The representation is a linear equation that combines a specific set of input values x the solution to which is the predicted output for that set of input values y. As such, both the input values x and the output value are numeric.

5.Simple Linear Regression

- Simple linear regression is a linear regression model with a single independent variable
- Model for single dimension

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \tag{1}$$

- Naming:
  - The unknown parameters - β
  - The independent variables - X or x
  - The dependent variable - Y or y
  - Introduced error - ϵ

5.Multiple dimension extension.

- When there is a single input variable x, the method is referred to as simple linear regression. When there are multiple input variables, literature from statistics often refers to the method as multiple linear regression.
- Model for n dimension

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + ... + \beta_n x_{i,n} + \epsilon_i \tag{2}$$

- To matrix representation

$$y_i = x_i^T + \epsilon_i \tag{3}$$

$$Y = X\beta + \epsilon \tag{4}$$

5.Ordinary least squares

Method for estimating the unknown parameters in a linear regression model

$$\hat{\beta} = argmin_{\beta} S(\beta) \tag{5}$$

$$S(\beta) = \sum_{i=1}^{n} \mid y_i - \sum_{j=1}^{p} X_{ij}\beta_j \mid^2 \tag{6}$$

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_{2i} & \cdots & \sum_{i=1}^{n} x_{ki} \\ \sum_{i=1}^{n} x_{2i} & \sum_{i=1}^{n} x_{2i}^2 & \cdots & \sum_{i=1}^{n} x_{2i}x_{ki} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} x_{ki} & \sum_{i=1}^{n} x_{ki}x_{2i} & \cdots & \sum_{i=1}^{n} x_{ki}^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{2i}y_i \\ \vdots \\ \sum_{i=1}^{n} x_{ki}y_i \end{bmatrix}$$

5.Ordinary least squares (cont.)

Which leads to:

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = \hat{\boldsymbol{\beta}} = [\mathbf{X}'\mathbf{X}]^{-1}\,\mathbf{X}'\mathbf{y}$$

Gradient descent

Gradient descent is a first-order iterative optimization algorithm for finding the minimum of a function.
Denote E as squared mean error in example for linear regression:

$$E = 1/n * \sum_{i=0}^{n} (y_i - \beta_1 * x_i + \beta_0) \tag{7}$$

Lets assume that:

$$a = \beta_1 \tag{8}$$

$$b = \beta_0 \tag{9}$$

$$D_a = \frac{\mathrm{d}}{\mathrm{d}a}(E) \tag{10}$$

$$D_b = \frac{\mathrm{d}}{\mathrm{d}b}(E) \tag{11}$$

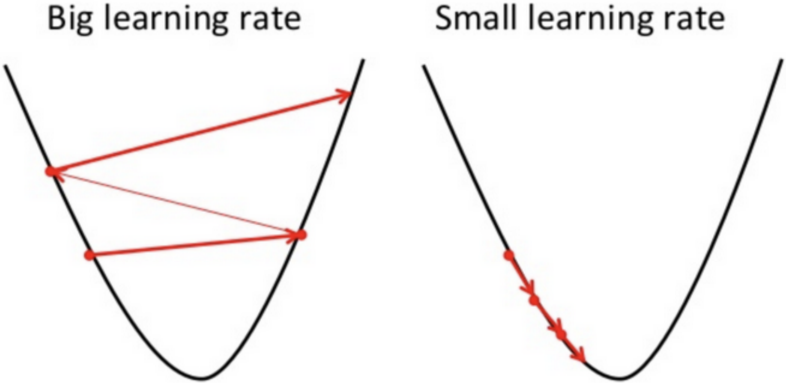
Gradient descent (cont.)

Now find that a and b where function E will reach minimum or be small enough. Take some L as learning rate and iterative find values of a and b from equations:

$$a = a - L * D_a \tag{12}$$

$$b = b - L * D_b \tag{13}$$

Gradient descent (cont.)



Training rate L has to be small enough to avoid skipping minimum.

Regularization

6. Logistic regression

7.Polynominal Regression

- Occurs when regression equation has independent variable in power higher than 1.
- General equation

$$y = \beta_0 + \beta_1 * x_i + \beta_2 * x_i^2 + \beta_3 * x_i^3 + ... + \beta_m * x_i^m + E, i = 1, 2, 3... \tag{14}$$

- Example (variable in 2 power)

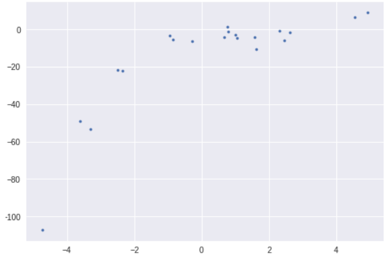
$$y = a + b * x^2 \tag{15}$$

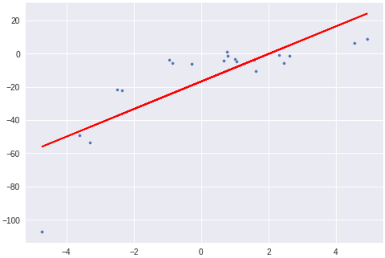
- the best fit is rather curve not a straight line

7.Polynominal Regression

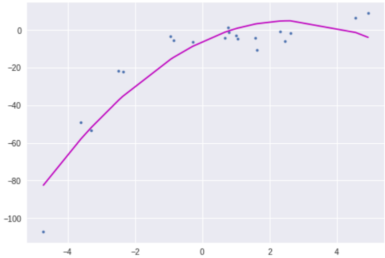
- polynominal with higher degree can give us lower error rate
- if degree will be too high then overfitting will occur
- curve sholud fit the nature of the problem (trend) not every single sample
- nonlinear "relationship" of variables, but it is considered as linear model due to the coefficients/weights associated with the features are still linear.

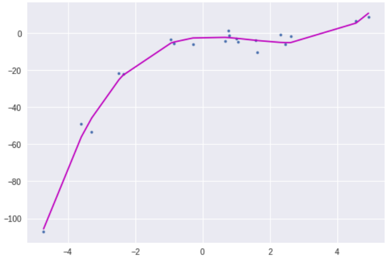
7.Polynominal Regression



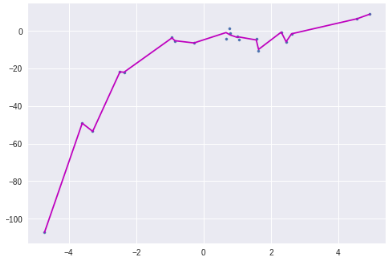


7.Polynominal Regression





7.Polynominal Regression



Thank you for your attention!

Q & A