PROGRAMMING FOR MACHINE LEARNING (D0036E)

Regression analysis

LULEÅ
UNIVERSITY
OF TECHNOLOGY

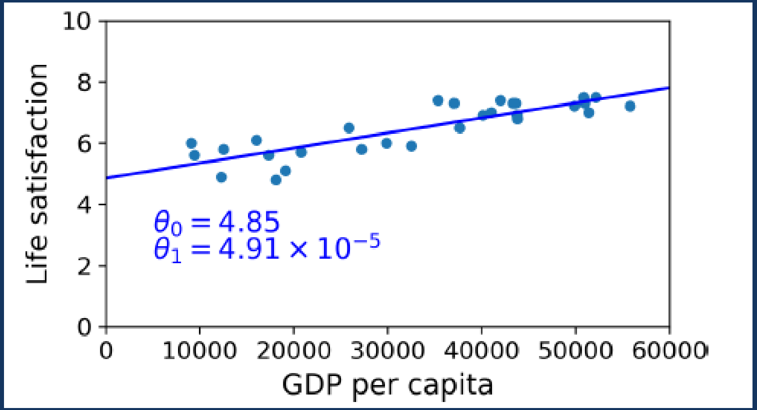
Regression analysis

As part of statistical analysis:

”Regression analysis is a set of statistical methods used for the estimation of relationships between a dependent variable and one or more independent variables”

The dependent variable is also known as outcome, response or label, predicted value in ML.

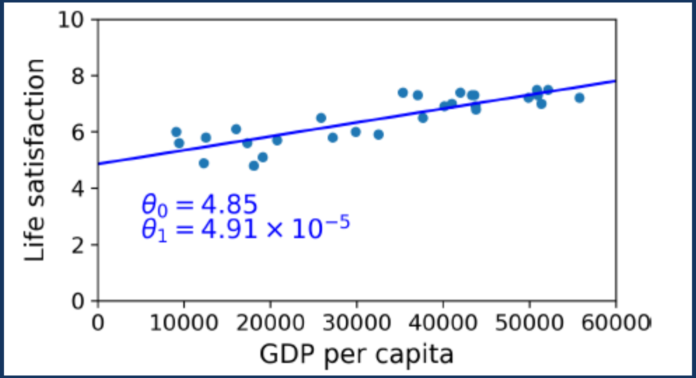
The independent variables are also known as explanatory variables, covariates or features in ML



Regression analysis

Uses of regression analysis

- Prediction
- Forecasting
- Inferring causal relationship



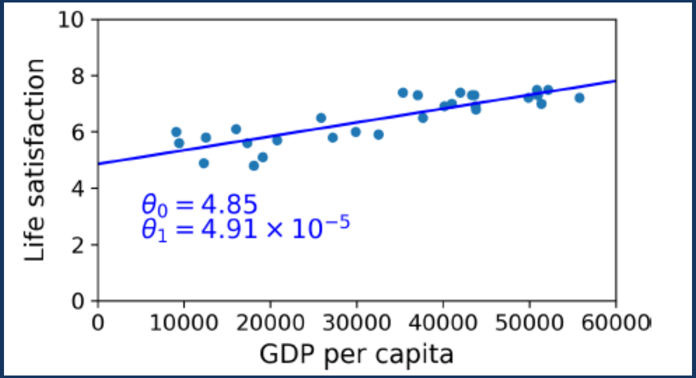
Types of regression analysis

- Linear regression
- Multivariate linear regression
- Non-linear regression
- And so on…

Regression analysis

Limits of regression analysis

1. Sample is representative for the population
2. (Lack of) A priori knowledge of function
3. Independent variables can’t be random
4. There is no (consistent) error on independent variables
5. Limited noise on independent variables
6. Helps if independent variables are independent from each other



Linear (multivariate) regression

Linear regression is a model that assesses the relationship between the predicted (dependent) variable and the feature(s) (independent variables).

Equation 4-1. Linear Regression model prediction

$$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$$

- $\hat{y}$ is the predicted value
- n is the number of features
- x_i is the i^{th} feature
- θ_i is the i^{th} model parameter

$\hat{y}$ is a linear combination of the features

Can be considered as the weighted sum of the input features

$(Y = a + b_1 x_1 + b_2 x_2 + \ldots \varepsilon)$, ε is the residual (error)

Linear (multivariate) regression

In vectorized form

Equation 4-2. Linear Regression model prediction (vectorized form)

$$\hat{y} = h_{\boldsymbol{\theta}}(\mathbf{x}) = \boldsymbol{\theta} \cdot \mathbf{x}$$

- $\hat{y}$ is the predicted value
- n is the number of features
- $\boldsymbol{\theta}$ is the parameter vector, including the bias term θ_0 and the feature weights $\theta_1, \dots, \theta_n$
- $\boldsymbol{x}$ is the feature vector, where $x_0 = 1$
- $\boldsymbol{\theta} \cdot \boldsymbol{x}$ is the dot product of $\boldsymbol{\theta}$ and $\boldsymbol{x}$
- h_θ is the hypothesis function using model parameters $\boldsymbol{\theta}$

In vector notation $\hat{y} = \boldsymbol{\theta}^T \boldsymbol{x}$, where $\boldsymbol{x}$ and $\boldsymbol{\theta}$ are column vectors

Linear (multivariate) regression

How do we measure the goodness of our model?

Equation 4-3. MSE cost function for a Linear Regression model

$$\mathrm{MSE}(\mathbf{X}, h_{\boldsymbol{\theta}}) = \frac{1}{m}\sum_{i=1}^{m}\left(\boldsymbol{\theta}^T\mathbf{x}^{(i)} - y^{(i)}\right)^2$$

Closed form solution for linear regression

Equation 4-4. Normal Equation

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \ \mathbf{X}^T \ \mathbf{y}$$

$\hat{\theta}$ is the value of θ that minimizes the cost function

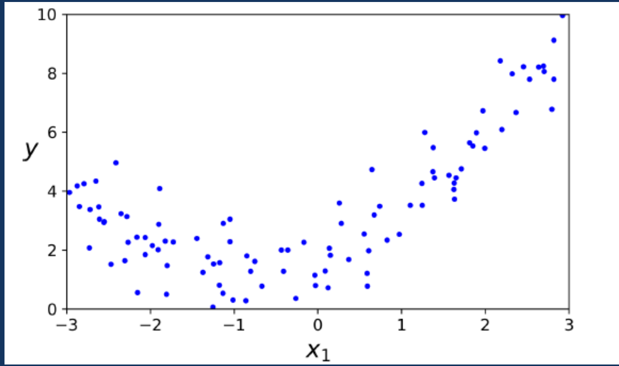
Matrix inversion is costly $O(n^3)$, so in practice a pseudoinverse can be used, which is $O(n^{2.4})$ (in scikit SVD is used)

There are other solution methods ….

Polynomial regression

Polynomial regression models the relationship between the predicted value (dependent variable) and the features (independent variables) as an n^{th} degree polynomial.

$$\hat{y} = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \ldots + \theta_i x^n$$
$$(\hat{y}_i = \theta_0 + \theta_1 x_i + \theta_2 x_i^2 + \theta_3 x_i^3 + \ldots + \theta_i x_i^n), i = 1 \, to \, m$$



Still a linear combination of values, so

Equation 4-4. Normal Equation

$$\hat{\theta} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{y}$$

applies.

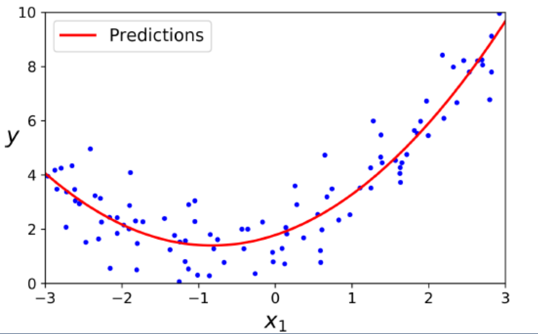
($\boldsymbol{X}$ is a Vandermonde matrix, invertible)

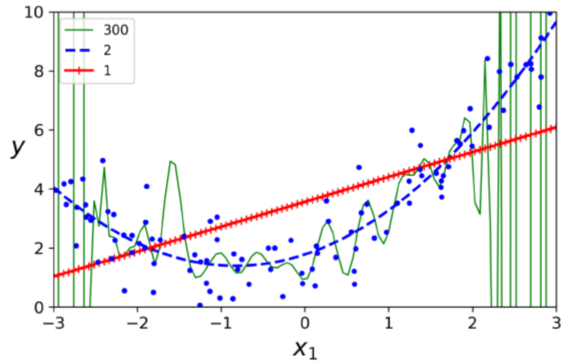
Polynomial regression

Limits of polynomial regression

(in addition to previously mentioned limits)

1. Not knowing the degree leads to overfitting or underfitting
2. Difficult to interpret parameters
3. “Weird” correlations





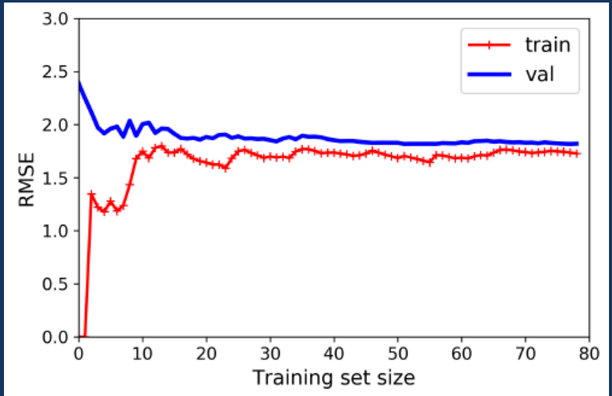
Polynomial regression

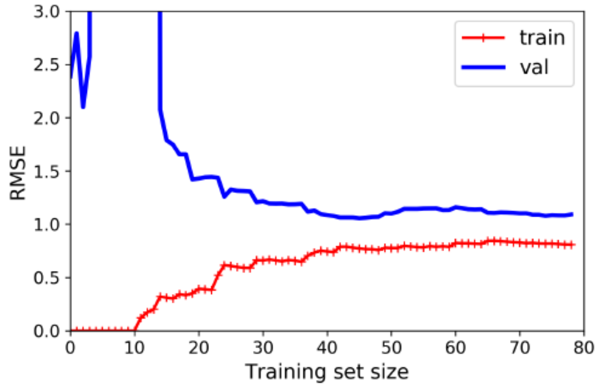
How do we know if we are over- or underfitting?

(Cross-validation)

Error is the sum of three factors:

1. Bias: Wrong assumptions lead to underfitting
2. Variance: Sensitivity to small fluctuations leads to overfitting
3. Irreducible error: Inherent noise, or measurement error





Non-linear regression

In non-linear regression, the relationship between the dependent and independent variables is modeled by an (arbitrary) function with a nonlinear combination of parameters.

E.g.:

$$\hat{y} = \frac{\theta_1 x}{\theta_2 + x}, \hat{y} = \theta_1 x^{\theta_2}, \ldots.$$

Non-trivial to fit.

There are solution methods…

Logistic regression

”Logistic regression estimates the probability of an event occurring, based on a given dataset of independent variables. Since the outcome is a probability, the dependent variable is bounded between 0 and 1.”

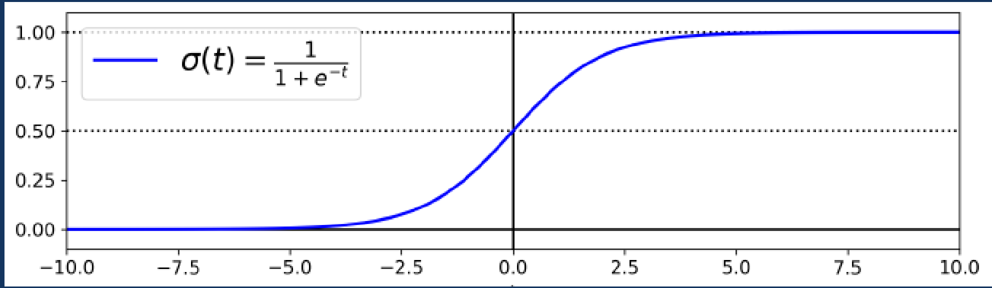
$$\hat{p} = h_\theta(x) = \sigma(\theta^T x)$$

Still a linear combination of features, but what is sigma?

Equation 4-14. Logistic function

$$\sigma(t) = \frac{1}{1 + \exp(-t)}$$

Logistic regression



$\sigma(t) < 0.5$ when $t < 0$, and $\sigma(t) \geq 0.5$ when $t \geq 0$, so a Logistic regression model predicts 1 if $\boldsymbol{\theta}^T \boldsymbol{x}$ is positive and 0 if it is negative

Equation 4-15. Logistic Regression model prediction

$$\hat{y} = \begin{cases} 0 & \text{if } \hat{p} < 0.5 \\ 1 & \text{if } \hat{p} \geq 0.5 \end{cases}$$

$(\hat{p} = h_\theta(x) = \sigma(\theta^T x))$

Figures and equations are from the course book

Géron, Aurélien. "Hands-on machine learning with scikit-learn and tensorflow: Concepts." Tools, and Techniques to build intelligent systems, O'Reilly (2017).

LULEÅ
UNIVERSITY
OF TECHNOLOGY