Machine Learning and Artificial Intelligence

Lab 08 – Linear and Polynomial Regression

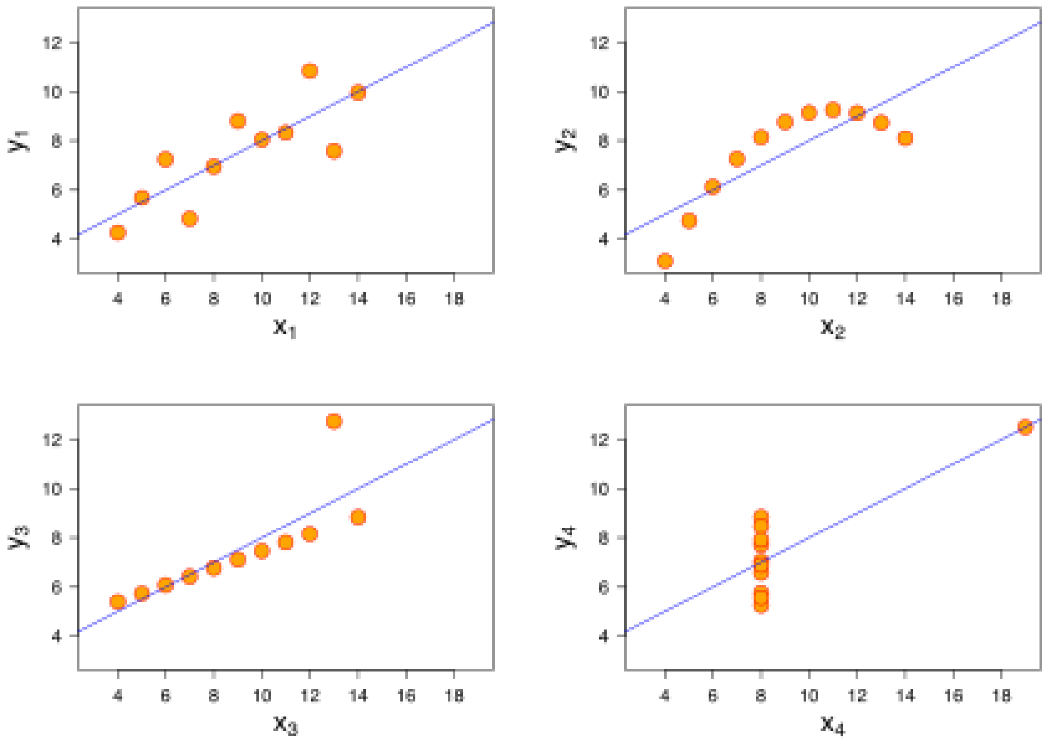
26/04/2021

Regression analysis

- Regression analysis is a set of statistical processes for estimating the relationships between a **dependent variable** and one or more **independent variables**.

- In these problems, the dependent variable is a continuous quantity, rather than a categorical one, which is what we have seen so far in our classification problems.

- Serves two main purposes:
  1. *Prediction and forecasting*
  2. ***(Pseudo)causal** inference*


y1
x1
y2
x2
y3
x3
y4
x4

Components

Regression models involve the following components:

- The **unknown parameters**.
- The **independent variables**, which are observed in data and are often represented by a vector.
- The **dependent variable**, which is observed in data and is often denoted using a scalar.
- The **error terms**, which are *not* directly observed in data and are often denoted using the scalar.

Linear regression

- One of the simplest and most widely used techniques for regression.
- Given $\{y_i, x_{i1}, \ldots, x_{ip}\}_{i=1}^{n}$, fits a line to the data:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i = \mathbf{x}_i^{\mathsf{T}} \boldsymbol{\beta} + \varepsilon_i, \qquad i = 1, \ldots, n,$$

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

- The error is represented by the *total* sum of squared errors:

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (y_i - \widehat{y_i})^2$$

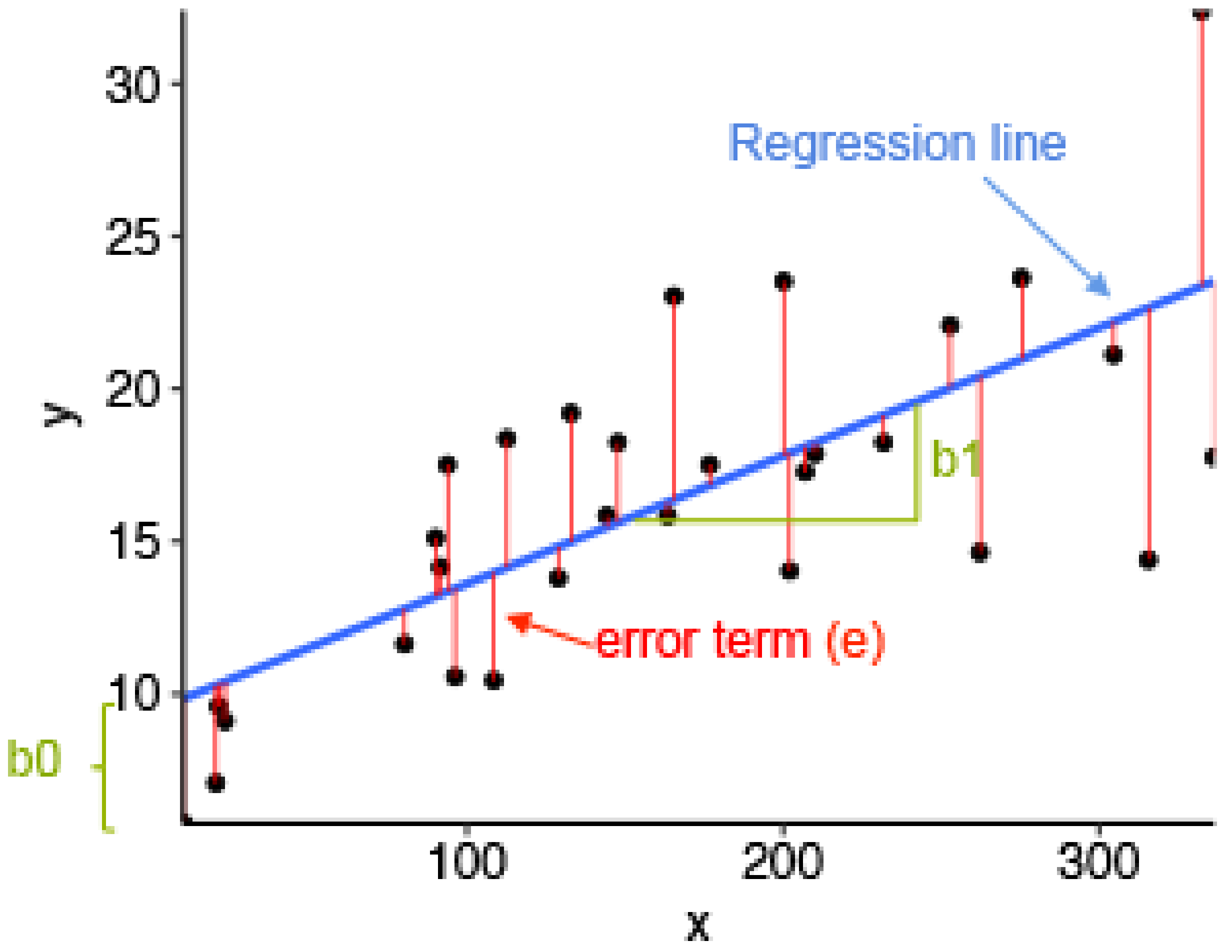
- **Fit the line in such a manner that this error is minimal!**

Least squares estimation

- One of the most optimal attributes of linear regression is that is **has a closed form solution**, given by ordinary *least squares*:

$$\mathrm{X} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix}, \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}, \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

$$\hat{\boldsymbol{\beta}} = \left(\mathrm{X}^{\mathsf{T}}\mathrm{X}\right)^{-1}\mathrm{X}^{\mathsf{T}}\mathbf{y}.$$

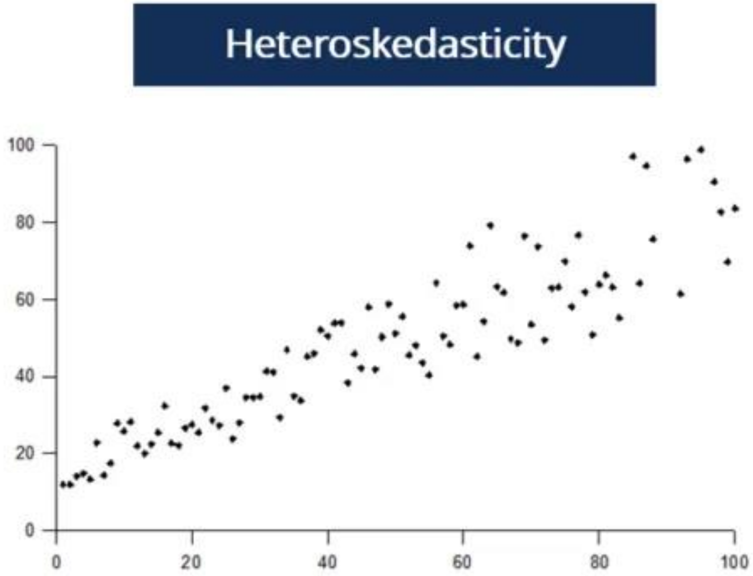


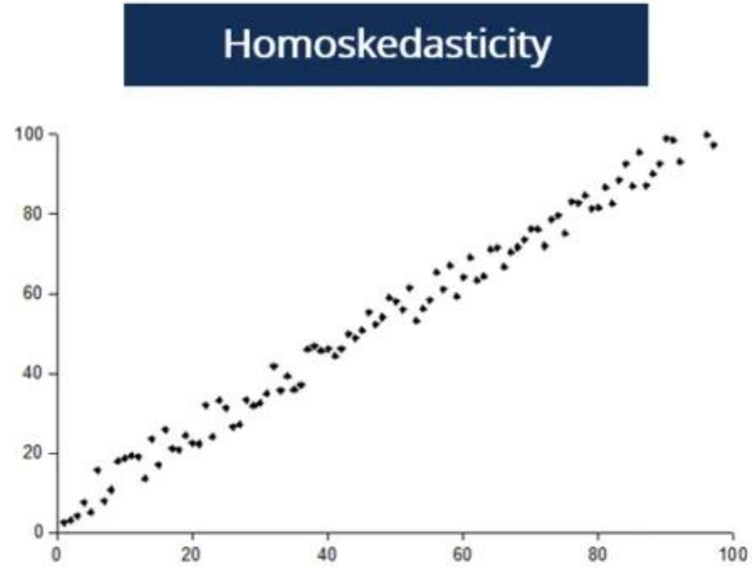
Regression line
b1
error term (e)
b0
30
25
20
15
10
y
100
200
300
x

Assumptions (a whole lot)

- The ordinary least squares solution makes a lot of assumptions in order to guarantee that the line is the one of best fit (least squared error):
  1. Linearity: The mean of the dependent variable is a linear combination of the independent variables (regressors).
  2. No linear dependence between regressors.
  3. Errors (Residuals) are independent and follow a normal distribution centered at 0 ($\mu = 0$)
  4. Homoskedasticity: Different values of the dependent variable have the same variance in their errors, regardless of the values of the predictor (independent) variables.

Assumptions





Polynomial regression

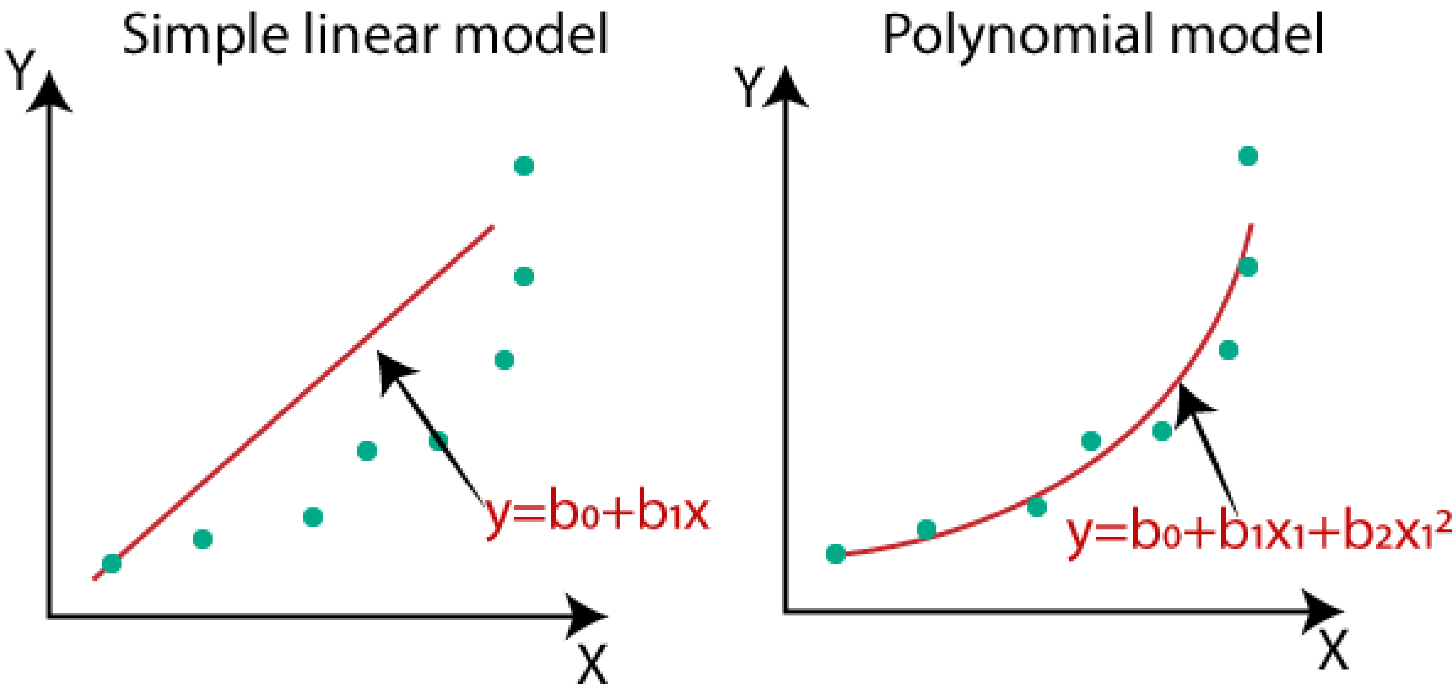
- Sometimes the relationship between the dependent and independent variables may not be linear.
- The relationship between the independent variable *x* and the dependent variable *y* can be modelled as an *n*th degree polynomial in *x.*

$$\boldsymbol{y_i \;=\; \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_m x_i^m + \varepsilon_i \;(i = 1, 2, \ldots, n)}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^m \\ 1 & x_2 & x_2^2 & \dots & x_2^m \\ 1 & x_3 & x_3^2 & \dots & x_3^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^m \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix},$$

- Fit the line in such a manner that SSE is minimal! (again)

Polynomial regression



OLS still holds

$$\hat{\boldsymbol{\beta}} = (\mathrm{X}^\mathsf{T}\mathrm{X})^{-1}\mathrm{X}^\mathsf{T}\mathbf{y}.$$

Goodness-of-fit

- In general, a model fits the data well if the differences between the observed values and the model's predicted values are small and unbiased.
- The most widely used statistic to test for a good model fit is R^2: it is the percentage of the response variable variation that is explained by a linear model.

$$R^2 = 1 - \frac{RSS}{TSS}$$

R^2 = coefficient of determination

RSS = sum of squares of residuals

TSS = total sum of squares

$$Adjusted\ R^2 = 1 - \frac{(1 - R^2)(N - 1)}{N - p - 1}$$

Other solutions?

- We can use Gradient Descent to solve the system as an optimization problem
- Imagine a Perceptron with no activation function and a squared error cost function.

$$J(\Theta_0, \Theta_1) = \frac{1}{2N}\sum_{i=1}^{N}(y_i - (\Theta_0 + \Theta_1 x_i))^2$$

$$\frac{\partial J}{\partial \Theta_0} = -\frac{1}{N}\sum_{i=1}^{N}(y_i - (\Theta_0 + \Theta_1 x_i))$$

$$\frac{\partial J}{\partial \Theta_1} = -\frac{1}{N}\sum_{i=1}^{N}x_i(y_i - (\Theta_0 + \Theta_1 x_i))$$

Exercises

Error metrics

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right|$$

(Other) Error metrics

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

$$\boldsymbol{M = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|}$$

Sklearn links

- https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html
- https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.PolynomialFeatures.html
- https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.SGDRegressor.html#sklearn.linear_model.SGDRegressor
- https://scikit-learn.org/stable/modules/generated/sklearn.metrics.r2_score.html#sklearn.metrics.r2_score
- https://scikit-learn.org/stable/modules/generated/sklearn.metrics.mean_absolute_error.html#sklearn.metrics.mean_absolute_error

Exercises