Introduction to task 1

I | On this list we look at linear regression from the optimisation point of view.

We start by denoting equation that expresses relationship between predictors (x_i) and outcomes (y_i)

$$\boxed{y_i = x_i \beta_0 + \varepsilon_i}$$

- y_i - denoted variable
- x_i - vector of regressors $[1 \times K]$
- β_0 - vector of regression coefficients
- ε_i - unobservable error term

Example:
We are estimating impact of different factors (year of production, KM, number of windows, brand) on price. $price = factors \cdot \beta + \varepsilon$

β_0 is vector of values that we are trying to estimate.

II | Matrix representation

For N observations, we can write this equation as

$y = X\beta_0 + \varepsilon$ where we add another (N) dimention to each part of the equation

III | Assumptions about error term

We make some assumptions about ε (normality and independence)

- Mean is 0
- Variance is $\sigma_0^2 I$, where I is an identity matrix (so we have single variable representing variance that is constant for all errors in ε term)
- Entries of ε are independent: $Cov(\varepsilon_i, \varepsilon_j) = 0,\ i \neq j$

IV | Cumulative Distribution Function

$$f_y(y_i | X) = (2\pi\sigma_0^2)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\frac{(y_i - x_i\beta_0)^2}{\sigma_0^2}\right)$$

This equation is expression of conditional distribition of y_i given X matrix.

V | Maximum Likelihood Estimate

We use MLE to estimate β_0 and σ_0^2. $\Rightarrow$ We aim to find values that maximize the likelyhood of observing the given data.

The likelihood function for entire dataset is product of the individual likelyhoods for each observation: $L(\beta_0, \sigma_0^2 | y, X) = \prod_{i=1}^{N} f_y(y_i | X)$

Taking the natural logarithm of the likelihood function (log-likelihood):

each observation. $L(\beta_0, \sigma_0 | y, X) = \prod_{i=1} f_Y(y_i | X)$

- Taking the natural logarithm of the likelihood function (log-likelyhood):

$$\log L(\beta_0, \sigma_0^2 | y, X) = -\frac{N}{2} \ln(2\pi\sigma_0^2) - \frac{1}{2\sigma_0^2} \sum_{i=1}^{N} (y_i - x_i\beta_0)^2$$

- Which can be used to estimate β_0 and σ_0^2 parameters.