In Class Exercises for 2[nd] week tutorials These are some in class problems repeating material from the Pre-Course. We will solve them in class in small groups and then discuss the solutions.

1. Linear Algebra

In the lecture, the optimal parameters $(\hat{\beta}_0, \hat{\beta}_1)$ of a linear regression model $\hat{y}(x) = \beta_0 + \beta_1 x$ were given by the formulas

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \qquad \hat{\beta}_1 = \frac{\sum_{n=1}^{N} (x_n - \bar{x})(y_n - \bar{y})}{\sum_{n=1}^{N} (x_n - \bar{x})^2}$$

show that these are equivalent to the following formulas by vectorizing the sums.

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \qquad \hat{\beta}_1 = \frac{\frac{1}{N} x^\mathsf{T} y - \bar{x} \cdot \bar{y}}{\frac{1}{N} x^\mathsf{T} x - \bar{x}^2}$$

2. Calculus

In the lecture the residual square error (RSS) of a model $\hat{y}(x, \beta)$ was defined as

$$\mathrm{RSS}(\beta) = \sum_{n=1}^{N} (y_n - \hat{y}(x_n))^2$$

In the case of a linear model $\hat{y}(x) = \beta_0 + \beta_1 x$, convince yourself that this is equal to

$$\mathrm{RSS}(\beta) = \|y - \hat{y}(x)\|_2^2 = \|y - \begin{pmatrix} 1, & x \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}\|_2^2 = \|y - \tilde{X}\beta\|_2^2$$

where $\tilde{X} = \begin{pmatrix} 1 & x \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$. Then compute the gradient $\nabla_\beta \, \mathrm{RSS}$ two times in two different ways: once using the summation notation and computing the partial derivatives $\frac{\partial}{\partial \beta_0} \mathrm{RSS}$ and $\frac{\partial}{\partial \beta_1} \mathrm{RSS}$ individually, and once via the vector notation computing both simultaneously.

3. Optimization

Use the gradient obtained from the last exercise to perform 2 steps of gradient descent

$$\beta^{(t+1)} = \beta^{(t)} - \eta \nabla_\beta \, \mathrm{RSS}$$

on the dataset (1). Start with $\beta^{(0)} = (1, 0)$ and use a learning rate of $\eta = 0.01$. Make a scatter-plot of the data and sketch $\hat{y}$ after each iteration.

x	y
1	1
2	3
4	3

4. Probability – Bayes Theorem

(Example 2.2.3.1 from Murphy's book). Consider taking a medical test that can detect breast-cancer. The test has a **sensitivity** of 80%, i.e. if the person has breast-cancer, then the test will be positive with a chance of 80%. Moreover, the test is also 90% **specific**, i.e. if a person does not have breast-cancer, the test will be negative with a chance of 90%.

A. Given that 0.4% of the total population suffer from breast cancer, how high is the probability that a person has cancer if the test result is positive?

B. How high is the chance that a person has cancer if they get tested positively two times with a different, independent tests (with same sensitivity and specificity)? How high after 3, 4 and 5 positive tests?