IN5400
Theoretical Exercises

Mohamed Ismail

April 2, 2020

These exercises can give you a hint about how exercises for the written exam can be.

1 Week 2

1.1 Linear Regression

a) What is the loss function for linear regression? Describe by words and formula

The loss function most commonly used in linear regression is the mean square error often shortened to MSE. It tells us the mean square error between the true y and the predicted $\hat{y}$ value summed over the m samples in the training set.

$$J(a,b) = MSE = \frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right)^2 = \frac{1}{m}\sum_{i=1}^{m}\left((ax^{(i)} + b) - y^{(i)}\right)^2 \tag{1}$$

b) How does the gradient descent algorithm update a and b?

The basic idea of gradient descent is that a function $F(\boldsymbol{x})$ decreases fastest if one goes from $\boldsymbol{x}$ in the direction of the negative gradient $\nabla F(\boldsymbol{x})$.

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \lambda \nabla F(\boldsymbol{x}_k) \tag{2}$$

$$a_{k+1} = a_k - \lambda \nabla J(a_k, b_k) \tag{3}$$

$$b_{k+1} = b_k - \lambda \nabla J(a_k, b_k) \tag{4}$$

λ is the learning rate, and for small enough λ then $F(\boldsymbol{x}_{k+1}) \leq F(\boldsymbol{x}_k)$. This means that for a sufficiently small λ we are always moving towards smaller function values, i.e a minimum. If λ is too large, the algorithm may diverge, and if it is too small, then the algorithm may converge slowly.

1.2 Cost function & Gradient Descent

You are given a vector a measurements $x^{(i)}$ and true values $y^{(i)}$

$$x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad y = \begin{bmatrix} 1.5 \\ 2 \\ 2.5 \end{bmatrix} \tag{5}$$

a) Plot y and x as points.

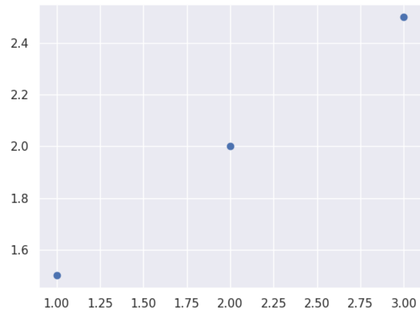

Figure 1: Caption

b) If we start with $w = 0$ and $b = 0$, what is the initial value for the loss function?

$$\begin{aligned} J(a,b) = MSE = \frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right)^2 &= \frac{1}{m}\sum_{i=1}^{m}\left((ax^{(i)} + b) - y^{(i)}\right)^2 \\ &= \frac{1}{m}\sum_{i=1}^{m}\left((0 * x^{(i)} + 0) - y^{(i)}\right)^2 \\ &= \frac{1}{m}\sum_{i=1}^{m}\left(- y^{(i)}\right)^2 = \frac{1}{3}\left(1.5^2 + 2^2 + 2.5^2\right)^2 = 2.083 \end{aligned}$$

c) Compute the next estimate of w and b, after 1 iteration of gradient descent.

$$\lambda = 1 \tag{6}$$

$$a_{k+1} = a_k - \lambda \nabla J(a_k, b_k) = \frac{2}{m}\sum_{i=1}^{m}\left((ax^{(i)} + b) - y^{(i)}\right)x^{(i)} = 4.33 \tag{7}$$

$$b_{k+1} = b_k - \lambda \nabla J(a_k, b_k) = \frac{2}{m}\sum_{i=1}^{m}\left((ax^{(i)} + b) - y^{(i)}\right) = 2.0 \tag{8}$$