Notebook

July 23, 2019

0.1 1: Fitting our Simple Model

Now that we have defined a simple linear model and loss function, let's begin working on fitting our model to the data.

0.1.1 Question 1

Let's confirm our visual findings for optimal $\hat{\theta}$.

Question 1a First, find the analytical solution for the optimal $\hat{\theta}$ for average squared loss. Write up your solution in the cell below using LaTeX.

Hint: notice that we now have $\mathbf{x}$ and $\mathbf{y}$ instead of x and y. This means that when writing the empirical risk function $R(\mathbf{x}, \mathbf{y}, \theta)$, you'll need to take the average of the squared losses for each y_i, $f_\theta(x_i)$ pair. For tips on getting started, see chapter 10 of the textbook (https://www.textbook.ds100.org/ch/10/modeling_loss_functions.html). Note that if you click "Open in DataHub", you can access the LaTeX source code of the book chapter, which you might find handy for typing up your work. Show your work, i.e. don't just write the answer.

$$\begin{aligned}
L(\theta, \mathbf{x}, \mathbf{y}) &= \frac{1}{n}\sum_{i=1}^{n} L(f_\theta(x_i), y_i) \\
&= \frac{1}{n}\sum_{i=1}^{n} L(\theta \cdot x_i, y_i) \\
&= \frac{1}{n}\sum_{i=1}^{n} (y_i - \theta \cdot x_i)^2
\end{aligned}$$

where

$$f_\theta(x_i) = \theta \cdot x_i$$

$$\begin{aligned}
\frac{\partial}{\partial \theta} L(\theta, \mathbf{x}, \mathbf{y}) &= \frac{-2}{n} \begin{bmatrix} (y_1 - \theta \cdot x_1)(x_1) \\ (y_2 - \theta \cdot x_2)(x_2) \\ \cdots \\ (y_i - \theta \cdot x_i)(x_i) \end{bmatrix} \\
&= \frac{-2}{n} (\mathbf{y} - \theta \cdot \mathbf{x})\mathbf{x}
\end{aligned}$$

In order to find the optimalθthat minimize the loss,

$$(\mathbf{y} - \theta \cdot \mathbf{x})\mathbf{x} = 0 \mathbf{y} \cdot \mathbf{x} - \theta \cdot (\mathbf{x})^2 = 0 \theta = \frac{\mathbf{y} \cdot \mathbf{x}}{(\mathbf{x})^2}$$

Therefore, the optimal$\hat{\theta}$is $\frac{\mathbf{y} \cdot \mathbf{x}}{(\mathbf{x})^2}$

Question 2a Use the average squared loss to compute $\frac{\partial R}{\partial \theta_1}, \frac{\partial R}{\partial \theta_2}$.

First, we will use LaTex to write $R(\mathbf{x}, \mathbf{y}, \theta_1, \theta_2)$, $\frac{\partial R}{\partial \theta_1}$, and $\frac{\partial R}{\partial \theta_2}$ given $\mathbf{x}$, $\mathbf{y}$, $\boldsymbol{\theta}$.

$$
\begin{aligned}
R(\mathbf{x}, \mathbf{y}, \theta_1, \theta_2) &= \frac{1}{n} \sum_{i=1}^{n} R(f_\theta(x_i), y_i) \\
&= \frac{1}{n} \sum_{i=1}^{n} (y_i - f_\theta(x_i))^2
\end{aligned}
$$

where$f_\theta(x_i) = \theta_1 \cdot x_i + sin(\theta_2 \cdot x_i)$

$$
\begin{aligned}
\frac{\partial}{\partial \theta_1} R = 2(y_i - \theta_1 \cdot x_i - sin(\theta_2 \cdot x_i))(-x_i) \\
\frac{\partial}{\partial \theta_2} R = 2(y_i - \theta_1 \cdot x_i - sin(\theta_2 \cdot x_i))(-x_i) cos(\theta_2 \cdot x_i)
\end{aligned}
$$