MLPM Tutorial 4

October 29, 2019

1. Given the data $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$ show that the line fit with ordinary least squares regression passes through the point $(\bar{x}, \bar{y})$ where $\bar{x} = \frac{1}{N}\sum_{i=1}^N x_i$ and $\bar{y} = \frac{1}{N}\sum_{i=1}^N y_i$.

2. Consider the regression model $p(y_i|\mathbf{x}_i) = \mathcal{N}(y_i|\hat{\mathbf{w}}_i^T \mathbf{x}_i, \sigma^2)$ where $\hat{\mathbf{w}}$ is the ordinary least squares estimate. Find the MLE estimate for σ^2.

3. Linear regression has the form $\mathbb{E}[y|\mathbf{x}] = w_0 + \mathbf{w}^T\mathbf{x}$. It is common to include a column of 1's in the design matrix, so we can solve for the offset term w_0 term and other parameters $\mathbf{w}$ at the same time using the normal equations. However, it is also possible to solve for $\mathbf{w}$ and w_0 separately. Show that
$$\hat{w_0} = \frac{1}{N}\sum_{i=1}^N y_i - \frac{1}{N}\sum_{i=1}^N \mathbf{w}^T\mathbf{x}_i = \bar{y} - \mathbf{w}^T\bar{\mathbf{x}}$$
So w_0 models the difference in the average output from the average predicted output. Also, show that
$$\hat{\mathbf{w}} = (\mathbf{X}_c^T\mathbf{X}_c)^{-1}\mathbf{X}_c^T\mathbf{y}_c = \left[\sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T\right]^{-1}\left[\sum_{i=1}^N (y_i - \bar{y})(\mathbf{x}_i - \bar{\mathbf{x}})\right]$$
where $\mathbf{X}_c$ is the centered input matrix containing $\mathbf{x}_i^c = \mathbf{x_i} - \bar{\mathbf{x}}$ along its rows and $\mathbf{y_c} = \mathbf{y} - \bar{\mathbf{y}}$ is the centered output vector. Thus we can first compute $\mathbf{w}$ on centered data, and then estimate w_0 using $\bar{y} - \hat{\mathbf{w}}^T\bar{\mathbf{x}}$.

4. Simple linear regression refers to the case where the input is scalar. Show that the MLE in this case is given by the following equations:
$$w_1 = \frac{\sum_{i=1}^N x_i y_i - N\bar{x}\bar{y}}{\sum_{i=1}^N x_i^2 - N\bar{x}^2} \approx \frac{\text{cov}[X, Y]}{\text{var}[X]}$$

$$w_0 = \bar{y} - w_1\bar{x} \approx \mathbb{E}[Y] - w_1\mathbb{E}[X]$$

5. The ordinary least squares estimate $\hat{\mathbf{w}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ requires access to the entire dataset $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$ all at once. Is it possible to find the same estimate $\hat{\mathbf{w}}$ in an online fashion where we see elements from $\mathcal{D}$ one by one? If so how?

6. Using your favourite programming language write an interactive program which recreates Figure 7.11 from the textbook. The user should be able to click on a plot in order to provide data points and the likelihood and posterior distributions should be visualised and updated with each new datapoint.