Notes On Linear Regression - Solutions to Exercises

CSE 446: Machine Learning
Autumn 2019

a. What is $\hat{w}_{MLE}$ under the Laplace distribution?

From the problem statement, we know that $\epsilon_i \sim Laplace(0, b)$, which implies that that $y_i = {x_i}^T w + \epsilon_i \sim Laplace({x_i}^T w, b)$.

Then, by the probability density function of the Laplace distribution, we know that

$$\begin{aligned} P(y_i|x_i, w, \epsilon_i) &= \frac{1}{2b} \exp\Bigg(- \frac{\left\| x^T w + \epsilon_i - x^T w \right\|_{L1}}{b} \Bigg) \\ &= \frac{1}{2b} \exp\Bigg(- \frac{\|\epsilon_i\|_{L1}}{b} \Bigg) \end{aligned}$$

Therefore, by the same maximum likelihood estimation logic we used in the normal distribution case, we know that

$$\begin{aligned} \hat{w}_{MLE} &= \text{argmax}_w \prod_{i=1}^{N} \Bigg[\frac{1}{2b} \exp\Bigg(- \frac{\|\epsilon_i\|_{L1}}{b} \Bigg) \Bigg] \\ &= \text{argmax}_w \sum_{i=1}^{N} \log \Bigg(\frac{1}{2b} \exp\Bigg(- \frac{\|\epsilon_i\|_{L1}}{b} \Bigg) \Bigg) \\ &= \text{argmax}_w \sum_{i=1}^{N} \Bigg(- \frac{\|\epsilon_i\|_{L1}}{b} \Bigg) \\ &= \text{argmin}_w \sum_{i=1}^{N} \Bigg(\frac{\|\epsilon_i\|_{L1}}{b} \Bigg) \\ &= \text{argmin}_w \sum_{i=1}^{N} \Big(\|\epsilon_i\|_{L1} \Big) \end{aligned}$$

b. Does $\hat{w}_{MLE}$ have an analytical (closed-form) solution?

No. We have to use methods such as gradient descent to solve this minimization problem.