Problem 9.1

Faisal Memon

June 26, 2024

1 Problem

Consider a model where the prior distribution over the parameters is a normal distribution with mean zero and variance σ_ϕ^2 so that

$$Pr(\phi) = \prod_{j=1}^{J} \text{Norm}_{\phi_j}[0, \sigma_\phi^2].$$

where j indexes the model parameters. We now maximise $\prod_{i=1}^{I} Pr(y_i|x_i, \phi)Pr(\phi)$. Show that the associated loss of this model is equivalent to L2 regularization.

2 Answer

Bayes Theorem says

$$Pr(H|E) = \frac{Pr(H) \cdot Pr(E|H)}{Pr(E)}$$

In our case

$$Pr(\phi|y) = \frac{Pr(\phi) \cdot Pr(y|\phi)}{Pr(y)}$$

The *maximum a posteriori* criterion is $\underset{\phi}{argmax} \prod_{i=1}^{I} Pr(y_i|x_i, \phi)Pr(\phi)$

We use a negative log likelihood function by minimising the negative log.

$$\underset{\phi}{argmin} - 1 \cdot \left(log \left(\sum_{i=1}^{I} Pr(y_{i|x_i,\phi}) \right) + log \left(\sum_{i=1}^{I} Pr(\phi) \right) \right)$$

Expanding out the $Pr(\phi)$

$$\text{Norm}_{\phi_j}[0, \sigma_\phi^2] = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\phi_j^2/2\sigma^2)$$

So that

$$log \left(\sum_{i=1}^{I} Pr(\phi) \right) = \sum_{j=1}^{J} -\phi_j^2/2\sigma^2 + constant$$

Re-arranging into the canonical form, this is

$$\phi = \underset{\phi}{argmin} \left[\sum_{i=1}^{I} l_i[x_i, y_i] + \lambda \sum_{j=1}^{J} \phi_j^2 \right]$$

This represents L2 regularization.