Bayes Linear Classification 03 Prediction & Conclusion

Chen Gong

06 November 2019

根据上一节中提到的 Inference，我们已经成功的推断出了 $p(w|Data)$ 的分布。表述如下所示：

$$p(w|X,Y) \sim \mathcal{N}(\mu_w, \Sigma_w) \tag{1}$$

其中，

$$\Sigma_w^{-1} = \sigma^{-2}X^TX + \Sigma_p^{-1} \qquad \mu_w = \sigma^{-2}A^{-1}X^TY \qquad \Sigma_w^{-1} = A \tag{2}$$

而我们的 Prediction 过程，可以被描述为，给定一个 x^* 如果计算得到 y^*。而我们的模型建立如下所示：

$$\begin{cases} f(x) = w^Tx = x^Tw \\ y = f(x) + \varepsilon \qquad\qquad \varepsilon \sim \mathcal{N}(0, \sigma^2) \end{cases} \tag{3}$$

1 Prediction

模型预测的第一步为，

$$f(x^*) = {x^*}^Tw \tag{4}$$

而在 Inference 部分，我们得到了 $p(w|Data) = \mathcal{N}(\mu_w, \Sigma_w)$。所以，我们可以推断出，

$$f(x^*) = {x^*}^Tw \sim \mathcal{N}({x^*}^T\mu_w, {x^*}^T\Sigma_w x^*) \tag{5}$$

那么公式 (5) 我们可以写作：

$$p(f(x^*)|Data, x^*) \sim \mathcal{N}({x^*}^T\mu_w, {x^*}^T\Sigma_w x^*) \tag{6}$$

又因为 $y = f(x) + \varepsilon$，所以

$$p(y^*|Data, x^*) \sim \mathcal{N}({x^*}^T\mu_w, {x^*}^T\Sigma_w x^* + \sigma^2) \tag{7}$$

那么计算到这里，我们的模型预测也算是完成了。

2 Conclusion

Data：$D = \{(x_i, y_i)\}_{i=1}^N$，其中 $x_i \in \mathbb{R}^p$，$y_i \in \mathbb{R}$。

Model：

$$\begin{cases} f(x) = w^TX = x^Tw \\ y = f(x) + \varepsilon \qquad\qquad \varepsilon \sim \mathcal{N}(0, \sigma^2) \end{cases} \tag{8}$$

Bayesian Method：w 不在是一个未知的常数，w 而是一个概率分布。贝叶斯线性分类可以被分成两个部分，Inference 和 Prediction。

1. Inference：$p(w|Data)$ 是一个 posterior 分布，假定 $p(w|Data) = \mathcal{N}(\mu_w, \Sigma_w) \propto likelihood \times prior$。这里使用了共轭的小技巧，得到 posterior 一定是一个 Gaussian Distribution。在这一步中，我们的关键是求出 μ_w 和 Σ_w。

2. Prediction：这个问题实际上也就是，给定一个 x^* 如果计算得到 y^*。我们可以描述为：

$$p(y^*|Data, x^*) = \int_w p(y^*|w, Data, x^*)p(w|Data, x^*)dw \tag{9}$$

又因为，y^* 只依赖于 w 和 x^*，不依赖于历史数据，所以 $p(y^*|w, Data, x^*) = p(y^*|w, x^*)$。并且，$w$ 的获得与 x^* 没有关系，所以 $p(w|Data)$。所以：

$$p(y^*|Data, x^*) = \int_w p(y^*|w, x^*)p(w|Data)dw = \mathbb{E}_{w\sim p(w|Data)}[p(y^*|w, x^*)] \tag{10}$$

之后通过自共轭特性不用计算积分即可得到服从的正态分布。