贝叶斯线性回归

 我们知道，线性回归当噪声为高斯分布的时候，最小二乘损失导出的结果相当于对概率模型应用 MLE，引入参数的先验时，先验分布是高斯分布，那么 MAP的结果相当于岭回归的正则化，如果先验是拉普拉斯分布，那么相当于 Lasso 的正则化。这两种方案都是点估计方法。我们希望利用贝叶斯方法来求解参数的后验分布。

线性回归的模型假设为：

$$
\begin{aligned}
f(x)&=w^Tx \quad (1)\\
y&=f(x)+\varepsilon \quad (2)\\
\varepsilon&\sim\mathcal{N}(0,\sigma^2) \quad (3)
\end{aligned}
$$

在贝叶斯方法中，需要解决推断和预测两个问题。

推断

引入高斯先验：

$$
p(w)=\mathcal{N}(0,\Sigma_p) \tag{4}
$$

对参数的后验分布进行推断：

$$
p(w|X,Y)=\frac{p(w,Y|X)}{p(Y|X)}=\frac{p(Y|w,X)p(w|X)}{\int p(Y|w,X)p(w|X)dw} \tag{5}
$$

分母和参数无关，由于 $p(w|X)=p(w)$，代入先验得到：

$$
p(w|X,Y)\propto\prod_{i=1}^N\mathcal{N}(y_i|w^Tx_i,\sigma^2)\cdot\mathcal{N}(0,\Sigma_p) \tag{6}
$$

高斯分布取高斯先验的共轭分布依然是高斯分布，于是可以得到后验分布也是一个高斯分布。第一项：

$$
\begin{aligned}
\prod_{i=1}^N\mathcal{N}(y_i|w^Tx_i,\sigma^2)&=\frac{1}{(2\pi)^{N/2}\sigma^N}\exp(-\frac{1}{2\sigma^2}\sum_{i=1}^N(y_i-w^Tx_i)^2)\\
&=\frac{1}{(2\pi)^{N/2}\sigma^N}\exp(-\frac{1}{2}(Y-Xw)^T(\sigma^{-2}\mathbb{I})(Y-Xw))\\
&=\mathcal{N}(Xw,\sigma^2\mathbb{I}) \quad (7)
\end{aligned}
$$

代入上面的式子：

$$
p(w|X,Y)\propto\exp(-\frac{1}{2\sigma^2}(Y-Xw)^T\sigma^{-2}\mathbb{I}(Y-Xw)-\frac{1}{2}w^T\Sigma_p^{-1}w) \tag{8}
$$

假定最后得到的高斯分布为：$\mathcal{N}(\mu_w,\Sigma_w)$。对于上面的分布，采用配方的方式来得到最终的分布，指数上面的二次项为：

$$
-\frac{1}{2\sigma^2}w^TX^TXw-\frac{1}{2}w^T\Sigma_p^{-1}w \tag{9}
$$

于是：

$$\Sigma_w^{-1}=\sigma^{-2}X^TX+\Sigma_p^{-1}=A \tag{10}$$

一次项：

$$\frac{1}{2\sigma^2}2Y^TXw=\sigma^{-2}Y^TXw \tag{11}$$

于是：

$$\mu_w^T\Sigma_w^{-1}=\sigma^{-2}Y^TX\Rightarrow\mu_w=\sigma^{-2}A^{-1}X^TY \tag{12}$$

预测

给定一个 x^*，求解 y^*，所以 $f(x^*)=x^{*T}w$，代入参数后验，有 $x^{*T}w\sim\mathcal{N}(x^{*T}\mu_w,x^{*T}\Sigma_wx^*)$，添上噪声项：

$$\begin{aligned}p(y^*|X,Y,x^*)=\int_wp(y^*|w,X,Y,x^*)p(w|X,Y,x^*)dw=\int_wp(y^*|w,x^*)p(w|X,Y)dw\\=\mathcal{N}(x^{*T}\mu_w,x^{*T}\Sigma_wx^*+\sigma^2)\end{aligned} \tag{13}$$