Ian Dardik
CS6140
Bayesian Approach to Linear Regression

1 MAP Estimation

We can build a discriminative model by maximizing $P(\theta|X,Y)$, i.e. by maximizing the likleihood that our model θ fit the given data X, Y. However, we can use a generative approach called Maximum a Posteriori to view the problem through a new lens. We begin by transforming the likelihood:

$$\begin{aligned} P(\theta|X,Y) &= \frac{P(\theta,X,Y)}{P(X,Y)} \\ &= \left(\frac{P(\theta,X,Y)}{P(\theta,X)}\right)\left(\frac{P(\theta,X)}{P(X)}\right)\left(\frac{P(X)}{P(X,Y)}\right) \\ &= \frac{P(Y|X,\theta)P(\theta|X)}{P(Y|X)} \end{aligned} \tag{1}$$

In many cases we can assume that [the data is fixed?] and therefore $P(\theta|X) = P(\theta)$ [2]. Therefore

$$P(\theta|X,Y) = \frac{P(Y|X,\theta)P(\theta)}{P(Y|X)}$$

We can maximize the log likleihood, notice that $P(Y|X)$ disapears because it's not dependent on θ:

$$\max_\theta \{\log P(Y|X,\theta) + \log P(\theta)\}$$

We assume assume our data is independent, and that $Y|X,\theta \sim \mathcal{N}(X'\theta, \sigma^2)$, and hence $P(Y|X,\theta) = \prod\limits_i \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{Y_i - X_i'\theta}{\sigma}\right)^2}$. We have many options for $P(\theta)$, we explore two below:

1. $\theta \sim \mathcal{N}(0, \frac{1}{\lambda}I)$, then:

$$\max_\theta \left\{\sum_i (Y_i - X_i'\theta)^2 + \lambda\theta'\theta\right\}$$

   Which is exactly the ridge solution.

2. $\theta \sim \text{Laplace}(0, \frac{1}{\lambda}I)$, then:

$$\max_\theta \left\{\sum_i (Y_i - X_i'\theta)^2 + \lambda \sum_{j=1}^{P} |\theta_j|\right\}$$

   Which is exactly the LASSO solution.

References

[1] Trevor Hastie, Robert Tibshirani, and Jerome Friedman, The Elements of Statistical Learning 2nd Edition

[2] Olga Vitek Generative Models, class slides