CS 438 Problem Set #2

Tony Melo

March 15, 2017

1.

(a)

$$L(y, f(X)) = \prod_{i=1}^{N} P(y^{(i)}|X^i)$$

Adding in the prior we get that:

$$L(y, f(X)) = \prod_{i=1}^{N} P(y^{(i)}|X^i) \prod_{i=1}^{k} N(\mu_p, \Sigma_p)$$

$$log(L(y, f(X))) = -log \sum_{i=1}^{N} P(y^{(i)}|X^i) log \sum_{i=1}^{k} P(\beta|\mu_p, \Sigma_p)$$

$$log(L(y, f(X))) = -log \sum_{i=1}^{N} \frac{1}{2\pi\sigma^2} - \frac{1}{2\sigma^2}(y^{(i)} - (\beta_o - \sum_{j=1}^{k} x_j^{(i)} \beta_j))^2 \cdot log \sum_{i=1}^{k} \frac{1}{2\pi(\Sigma_p)^2} - \frac{1}{2(\Sigma_p)^2}(\beta_k - \mu_p)^2$$

$$log(L(y, f(X))) = -\frac{1}{2} log \sum_{i=1}^{N} \frac{1}{\sigma^2}(y^{(i)} - (\beta_o - \sum_{j=1}^{k} x_j^{(i)} \beta_j))^2 \cdot log \sum_{i=1}^{k} \frac{1}{(\Sigma_p)^2}(\beta_k - \mu_p)^2$$

We see that this is exactly the ridge regression loss function if we maximize the above with respect to β and standardize the gaussian prior to have 0 mean and a variance of $\frac{1}{\lambda}$