Tutorial 2

Monday 16^{th} January, 2017

Problem 1. Posterior Distribution of w with very imprecise prior:
Let $y = \mathbf{w}^T\phi(\mathbf{x}) + \varepsilon$ and let dataset $\mathcal{D} = \{(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_i, Y_i), \ldots, (\mathbf{X}_m, Y_m)\}$ was provided. Recall that the posterior distribution for $\mathbf{w}$ under a Gaussian prior was $\Pr(\mathbf{w} \mid \mathcal{D}) = \mathcal{N}(\mathbf{w} \mid \mu_m, \Sigma_m)$ where

$$\Sigma_m^{-1} = \lambda I + \Phi^T\Phi/\sigma^2$$

and

$$\mu_m = (\lambda\sigma^2 I + \Phi^T\Phi)^{-1}\Phi^T\mathbf{y}$$

How would you model a very imprecise Gaussian prior on $\Pr(\mathbf{w})$? Explain what happens to the parameters of the posterior $\Pr(\mathbf{w} \mid \mathcal{D})$ as this precision on the prior $\Pr(\mathbf{w})$ tends to 0. What is the connection between this expression and the data likelihood expression?

Problem 2. Case for non-IID dataset:
In the class, we discussed the case of Bayesian estimation for a univariate Gaussian from dataset $\mathcal{D}$ that consisted of IID (independent and identically distributed) observations.

- Let $\Pr(X) \sim \mathcal{N}(\mu, \sigma^2)$ and let the data $\mathcal{D} = x_1 ... x_m$ be IID. Let σ^2 be known.
- $\mu_{MLE} = \frac{1}{m}\sum_{i=1}^{m} x_i$ and $\sigma_{MLE} = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu)^2$
- The conjugate prior is $\Pr(\mu) = \mathcal{N}(\mu_0, \sigma_0^2)$, And the **posterior** is: $\Pr(\mu | x_1 ... x_m) = \mathcal{N}(\mu_m, \sigma_m^2)$ such that
- $\mu_m = (\frac{\sigma^2}{m\sigma_0^2 + \sigma^2}\mu_0) + (\frac{m\sigma_0^2}{m\sigma_0^2 + \sigma^2}\hat{\mu}_{ML})$ and $\frac{1}{\sigma_m^2} = \frac{1}{\sigma_0^2} + \frac{m}{\sigma^2}$

Prove the above

Assume that $\Omega \in \Re^{m \times m}$ is positive-definite. Now answer the following questions

1. What would be the maximum likelihood estimate for μ?
2. How would you go about doing Bayesian estimation for μ?
3. What will be an appropriate conjugate prior?

4. What will the posterior be? And what will be the MAP and Bayes estimates?

   HOMEWORK: What about the special cases of Ω being diagonal matrices with the same or different values along the diagonal?

Problem 3. We discussed atleast two settings where maximizing a monotonically increasing function of the objective is somewhat more intuitive than maximizing the original objective. Recall the two settings. Now prove that maximizing the monotonically increasing transformation of the objective gives the same optimality point as does maximizing the original objective.