Tutorial 2

Thursday 4^{th} August, 2016

Problem 1. 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
- $\mu_m = (\dfrac{\sigma^2}{m\sigma_0^2 + \sigma^2}\mu_0) + (\dfrac{m\sigma_0^2}{m\sigma_0^2 + \sigma^2}\hat{\mu}_{ML})$ and $\dfrac{1}{\sigma_m^2} = \dfrac{1}{\sigma_0^2} + \dfrac{m}{\sigma^2}$

Prove the above. Now suppose, the examples $x_1 ... x_m$ in the dataset $\mathcal{D}$ were not necessarily independent and whose possible dependence was expressed by known covariance matrix Ω but with a common unknown (to be estimated) mean $\mu \in \Re$. Let $\mathbf{u} = [1, 1, \ldots 1]$ a $m-$dimensional vector of 1's and $\mathbf{x} = [x_1 ... x_m]$ and

$$Pr(x_1 ... x_m; \mu, \Omega) = \frac{1}{(2\pi)^{\frac{m}{2}}|\Omega|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\mu\mathbf{u})^T\Omega^{-1}(\mathbf{x}-\mu\mathbf{u})}$$

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?

Problem 2. 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.