COGS 118A: ASSIGNMENT 4

Problem 1. MLE, MAP, and Bayesian prediction
(a) Suppose we have observed data pairs \((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\), which are generated by the following relationship:
\[
y = w^T \phi(x) + \epsilon, \quad \text{where } \epsilon \sim \mathcal{N}(0, \sigma^2)
\]
and we also have a Gaussian prior distribution over the weights: \(w \sim \mathcal{N}(0, \tau^2 I)\). Show that the MAP estimate is the same as the Bayesian predictive mean (this is because the mode of a Gaussian distribution is also its mean).

(b) (Matlab) Generate 5 samples of \(x\) uniformly between -10 and 10. Assume \(\tilde{y} = x^2 - 4x - 21\), generate 5 corresponding samples of \(y\) from a Gaussian distribution \(\mathcal{N}(\tilde{y}, \sigma^2)\), where \(\sigma^2 = 20^2\). We will use a set of five polynomial basis functions: \(\phi_j(x) = x^{j-1}\), for \(j = 1, \ldots, 5\). Generate a set of new \(x\) values using the Matlab command \(x2 = (-10:.2:10)'\), find both the MLE (Bishop Eq. 3.15) and MAP (Bishop Eq. 3.55) estimates for all 101 new \(x\) values, where we assume the prior distribution for \(w\) is \(w \sim \mathcal{N}(0, \tau^2 I)\), where \(\tau = 15\). On the same figure, plot the true corresponding \(\tilde{y}\) for \(x2\) using a black line, plot the observations using green circles, plot the MLE for \(x2\) using a red line, plot the MAP estimate using a blue line. Print the figure. Compute the sum (over all values of \(x2\)) of the squared error between the estimated \(\tilde{y}\) and the true \(\tilde{y}\) for each of MLE and MAP estimates. What did you get? Repeat the process (starting from generating 5 samples of \(x\)) for 29 more times (but without generating the figure again), compute the mean and standard deviation of the summed square error for each of MLE and MAP.

Problem 2. Bias-Variance Trade-off versus Bayesian Model Comparison
(a) Suppose we have the likelihood function \(y \sim \mathcal{N}(w^T \phi(x), \sigma^2)\), and the prior, \(w \sim \mathcal{N}(0, \tau^2 I)\), we want to show that the model evidence (Bishop Eq. 3.68), is also a normal distribution
\[
p(y|M_r) = \int p(y|w, \sigma)p(w|\tau)dw = \mathcal{N}(0, \sigma^2 I + \tau^2 I^T) .
\]

Note that this is a distribution over \(y\), not over \(w\). It is related to, but simpler than, HW3 problem 3. Some hints about the steps of derivation:

- Write out the integral for \(p(y|M_r)\) and group like terms in the exponent.
• Use results from HW3 problem 3 to show that the function of $y$ (independent of $w$) we multiply in front of the exponential function, in order to complete the squares, is

$$f(y) = e^{-y^2(\frac{1}{2\sigma^2}I - \frac{1}{4\tau^2}S\phi^T)}$$

which is proportional to the marginal $p(y)$ (since the integral over $w$ is independent of $y$), where $S$ is the posterior covariance matrix.

• Show that this implies that $p(y)$ is also a normal distribution with mean 0.

• Use the Matrix Inversion Lemma (Eq. 10 in Sam Roweis’ notes on linear algebra) to show that the covariance matrix is $\sigma^2I + \tau^2\phi\phi^T$. This is a very useful lemma!

(c) (Matlab) Index models in (b) by the value of $\tau$, so that in $M_1$, $\tau = 4.5$, in $M_2$, $\tau = 9$, and in $M_3$, $\tau = 63$. What values of $\lambda$ do they correspond to? Use the 30 $(x, y)$ data points generated in (a) to compute the model evidence for each of the three models. What did you get for each? What are the three pair-wise Bayes factors? How did your results compare to the bias-variance analysis in (a)?