Report #2

1. *Maximum a posteriori estimator*. Suppose that we have n sample points (x^1, y^1), $\ldots$, $(x^n, y^n) \in \mathbf{R}^d \times \mathbf{R}$ with a linear prediction model in the following form

$$\hat{y}^i = \theta^T x^i$$

assuming the model responses being independently generated by

$$y^i \approx \theta^T x^i + \epsilon$$

where ϵ follows the zero-mean Gaussian distribution with variance σ^2. In other words:

$$p\left(y^i \mid \theta\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}\left(\theta^T x^i - y^i\right)^2\right)$$

We are interested in finding the *maximum a posteriori (MAP)* estimator, *i.,e., finding θ^* that maximizes the following posterior density function*,

$$p\left(\theta \mid y^1, \ldots, y^n\right)$$

The *prior* density, $p(\theta)$, tells us the likely values that θ may take *before* looking at the samples. Suppose that the elements of θ are independent zero-mean Gaussian with variance τ^2 following:

$$p(\theta) = \frac{1}{\left(2\pi\tau^2\right)^{d/2}} \exp\left(-\frac{1}{2\tau^2}\theta^T\theta\right)$$

We combine this with what the sample data tells us, namely, the *likelihood* density, $p(y^1, \ldots, y^n \mid \theta)$, using the Bayes' rule, and get the *posterior* density of θ, which tells us the likely θ value *after* looking at the samples. The Bayes' rule with the sample independence assumption leads to

$$\begin{aligned} p\left(\theta | y^1, \ldots, y^n\right) &= \frac{p\left(y^1, \ldots, y^n \mid \theta\right) p(\theta)}{\int p\left(y^1, \ldots, y^n \mid \eta\right) p(\eta)\, d\eta} \\ &= \alpha p(\theta) \prod_{i=1}^{n} p\left(y^i \mid \theta\right) \end{aligned}$$

for some $\alpha > 0$.

Now find the MAP estimator, that is, express the optimal θ^* in terms of your sample data points, (x^1, y^1), $\ldots$, (x^n, y^n). *Hint: taking logarithm of $p(\theta \mid y^1, \ldots, y^n)$ does not change your solution.*