Problem 1

Hand-in 1 Jonatan Haraldsson

Show that the MAP using the likelihood $p(d_i|\theta) = \mathcal{N}(d_i \,|\, [\Phi\theta]_i, \sigma^2)$ and the prior: $p(\theta) = \prod_{i=1}^{M_\theta} \mathcal{N}(\theta_i|0,\sigma_0^2)$

is equal to $\hat{\theta}_{Ridge} = \underset{\theta}{\operatorname{argmin}} \left\{ (\Phi\theta - D)^T(\Phi\theta - D) + \lambda \sum_{i=1}^{M_\theta} \theta_i^2 \right\}$

The likelihood is given by $p(D|\theta) = \mathcal{N}(D|\Phi\theta, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{N/2} \exp\left\{-\frac{1}{2}\frac{(D-\Phi\theta)^T(D-\Phi\theta)}{\sigma^2}\right\}$

Similarly the prior is given by: $p(\theta) = \prod_{i=1}^{M_\theta} \mathcal{N}(\theta_i|0,\sigma_0^2) = \prod_{i=1}^{M_\theta} \sqrt{\frac{1}{2\pi\sigma_0^2}} \exp\left\{-\frac{1}{2}\frac{\theta_i^2}{\sigma_0^2}\right\} = \left[\frac{1}{2\pi\sigma_0^2}\right]^{M_\theta/2} \exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{M_\theta}\theta_i^2\right\}$

From Bayes' the posterior $\propto$ "likelihood × prior". Thus, $\hat{\theta}_{MAP} = \underset{\theta}{\operatorname{argmax}} \{p(D|\theta)p(\theta)\}$.

Let's go!

$$p(D|\theta)p(\theta) = \left(\frac{1}{2\pi\sigma^2}\right)^{N/2} \exp\left\{-\frac{1}{2}\frac{(D-\Phi\theta)^T(D-\Phi\theta)}{\sigma^2}\right\} \left[\frac{1}{2\pi\sigma_0^2}\right]^{M_\theta/2} \exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{M_\theta}\theta_i^2\right\} \sim \left\{\begin{array}{l}\text{No } \theta \text{ dependence in the pre-factors,}\\ \text{so I'll only focus on the exponentials}\end{array}\right\} \sim$$

$$p(D|\theta)p(\theta) \sim \exp\left\{-\frac{1}{2}\frac{(D-\Phi\theta)^T(D-\Phi\theta)}{\sigma^2}\right\}\exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{M_\theta}\theta_i^2\right\} = \exp\left\{-\frac{1}{2\sigma^2}\left[(D-\Phi\theta)^T(D-\Phi\theta)\right] - \frac{1}{2\sigma_0^2}\sum_{i=1}^{M_\theta}\theta_i^2\right\} =$$

$$= \left\{\begin{array}{c}\text{Factor-out}\\ \frac{1}{2\sigma^2}\end{array}\right\} = \exp\Big\{-\frac{1}{2\sigma^2}\underbrace{\Big[(D-\Phi\theta)^T(D-\Phi\theta) + \frac{\sigma^2}{\sigma_0^2}\sum_{i=1}^{M_\theta}\theta_i^2\Big]}_{=f(\theta)}\Big\} = e^{-f(\theta)}$$

Finding $\underset{\theta}{\operatorname{argmax}}\{p(D|\theta)p(\theta)\} \Leftrightarrow \underset{\theta}{\operatorname{argmax}}\{e^{-f(\theta)}\} \Leftrightarrow \{e^{-f(\theta)} \text{ is max for } \min\{f(\theta)\}\} \Leftrightarrow \underset{\theta}{\operatorname{argmin}}\{f(\theta)\}$

To wrap up: $\hat{\theta}_{MAP} = \underset{\theta}{\operatorname{argmax}}\{p(D|\theta)p(\theta)\} = \underset{\theta}{\operatorname{argmin}}\{f(\theta)\} = \left\{\begin{array}{l}\text{Again, excluding}\\ \text{the } -\frac{1}{2\sigma^2} \text{ pre-factor}\end{array}\right\} = \underset{\theta}{\operatorname{argmin}}\left\{(D-\Phi\theta)^T(D-\Phi\theta) + \frac{\sigma^2}{\sigma_0^2}\sum_{i=1}^{M_\theta}\theta_i^2\right\}$

Comparing to where I'm supposed to arrive, $\lambda = \left(\frac{\sigma}{\sigma_0}\right)^2 \Rightarrow \hat{\theta} = \underset{\theta}{\operatorname{argmin}}\left\{(D-\Phi\theta)^T(D-\Phi\theta) + \lambda\sum_{i=1}^{M_\theta}\theta_i^2\right\}$.

Problem 2

Show that the MAP using the likelihood $p(d_i|\theta) = \mathcal{N}(d_i \,|\, [\Phi\theta]_i, \sigma^2)$ and the prior: $p(\theta) = \prod_{i=1}^{M_\theta} \mathcal{L}(\theta_i|0,\sigma_0)$

is equal to $\hat{\theta}_{Lasso} = \underset{\theta}{\operatorname{argmin}} \left\{ (\Phi\theta - D)^T(\Phi\theta - D) + \lambda \sum_{i=1}^{M_\theta} |\theta_i| \right\}$

I'll be following the same approach as in problem 1, i.e. finding $\underset{\theta}{\operatorname{argmax}}\{p(D|\theta)p(\theta)\}$

Course book Ch. 3.2, p.66

We've got the same likelihood, however, the prior is given by: $p(\theta) = \prod_{i=1}^{M_\theta}\mathcal{L}(\theta_i|0,\sigma_0)$, $\mathcal{L}(\theta_i|0,\sigma_0) \propto \exp\left\{-\frac{|\theta_i|}{\sigma_0}\right\}$

The prior is then: $p(\theta) \propto \prod_{i=1}^{M_\theta}\mathcal{L}(\theta_i|0,\sigma_0) = \prod_{i=1}^{M_\theta}\exp\left\{-\frac{|\theta_i|}{\sigma_0}\right\} = \exp\left\{-\frac{1}{\sigma_0}\sum_{i=1}^{M_\theta}|\theta_i|\right\}$

Again, omitting the θ-independent prefactors: $p(D|\theta)p(\theta) \propto \exp\left\{-\frac{1}{2}\frac{(D-\Phi\theta)^T(D-\Phi\theta)}{\sigma^2}\right\}\exp\left\{-\frac{1}{\sigma_0}\sum_{i=1}^{M_\theta}|\theta_i|\right\} =$

$$= \exp\left\{-\frac{1}{2\sigma^2}\left[(\Phi\theta-D)^T(\Phi\theta-D) + \frac{2\sigma^2}{\sigma_0}\sum_{i=1}^{M_\theta}|\theta_i|\right]\right\}$$

Yet again, $\underset{\theta}{\operatorname{argmax}}\left\{\exp\left\{-\frac{1}{2\sigma^2}\left[(\Phi\theta-D)^T(\Phi\theta-D) + \frac{2\sigma^2}{\sigma_0}\sum_{i=1}^{M_\theta}|\theta_i|\right]\right\}\right\} \Leftrightarrow \underset{\theta}{\operatorname{argmin}}\left\{(\Phi\theta-D)^T(\Phi\theta-D) + \frac{2\sigma^2}{\sigma_0}\sum_{i=1}^{M_\theta}|\theta_i|\right\}$

Comparing to what I want $\lambda = \frac{2\sigma^2}{\sigma_0}$ in this case.