Machine Learnig Homework 6

Problem 1

Let $\hat{\Phi} = \begin{pmatrix} \Phi \\ \sqrt{\lambda} I_{p \times M} \end{pmatrix} \in \mathrm{R}^{(N+p)\times M}$ be the augmented design matrix and $\hat{\mathbf{z}} = \begin{pmatrix} \mathbf{z} \\ \mathbf{0}_p \end{pmatrix} \in \mathrm{R}^{N+p}$ the augmented target vector. Maximum likelihood estimation of parameters $\mathbf{w}$ for linear regression is equivalent to minimizing energy given by $E(\mathbf{w}) = (\mathbf{z} - \Phi^T\mathbf{w})^T(\mathbf{z} - \Phi^T\mathbf{w})$. The minimizer $\mathbf{w}_{MLE} = (\Phi^T\Phi)^{-1}\Phi^T\mathbf{z}$ is agnostic of the form and contents of $\mathbf{w}$, $\mathbf{z}$ or Φ. Therefore, we can write the minimizer of the energy resulting from the augmented design matrix and target vector as

$$
\begin{aligned}
\hat{\mathbf{w}}_{MLE} &= (\hat{\Phi}^T\hat{\Phi})^{-1}\hat{\Phi}^T\hat{\mathbf{z}} && (1)\\
&= \left(\begin{pmatrix} \Phi^T & \sqrt{\lambda} I^T_{p\times M} \end{pmatrix} \begin{pmatrix} \Phi \\ \sqrt{\lambda} I_{p \times M} \end{pmatrix} \right)^{-1} \begin{pmatrix} \Phi^T & \sqrt{\lambda} I^T_{p\times M} \end{pmatrix} \begin{pmatrix} \mathbf{z} \\ \mathbf{0}_p \end{pmatrix} && (2)\\
&= \left(\lambda I_{M\times M} + \Phi^T\Phi\right)^{-1} \Phi^T\mathbf{z} && (3)\\
&= \mathbf{w}^{ridge}_{MLE} && (4)
\end{aligned}
$$

Problem 2

We want to prove the following identity:

$$(M + vv^T)^{-1} = M^{-1} - \frac{(M^{-1}v)(v^TM^{-1})}{1 + v^TM^{-1}v}$$

To prove the identity we will use the fact that for any matrix M:

$$MM^{-1} = I$$

where I is the identity matrix. So we multiply both sides by the inverse of the left-hand side

$$M + vv^T$$

which results in

$$I = I + M^{-1}vv^T - \frac{M^{-1}vv^T + M^{-1}vv^TM^{-1}vv^T}{1 + v^TM^{-1}v} \qquad (5)$$

$v^TM^{-1}v$ is a scalar, so:

$$M^{-1}vv^T + M^{-1}vv^TM^{-1}vv^T = (1 + v^TM^{-1}v)M^{-1}vv^T$$

(5) then simplifies to:

$$I = I + M^{-1}vv^T - M^{-1}vv^T = I$$

so the identity is verified.
Using the identity above, we need to prove the following inequality:

$$\sigma^2_{N+1}(x) \leq \sigma^2_N(x) \tag{6}$$

We use the definition of σ^2_N and the definition of S_N (here Φ_N is the design matrix with the first N observations).

$$\sigma^2_N(x) = \frac{1}{\beta} + \phi(x)^TS_N\phi(x)$$

$$S_N^{-1} = \alpha I + \beta(\Phi_N^T)\Phi_N$$

We now use the following identity:

$$(\Phi_N^T)\Phi_N = \sum_{k=1}^{N}\phi(x_k)\phi(x_k)^T$$

so we can write

$$(\phi_{N+1}^T)\phi_{N+1} = (\phi_N^T)\phi_N + \phi(x_{N+1})\phi(x_{N+1})^T$$

and hence

$$S_{N+1}^{-1} = \alpha I+\beta(\Phi_{N+1}^T)\Phi_{N+1} = \alpha I+\beta((\Phi_N^T)\Phi_N+\phi(x_{N+1})\phi(x_{N+1})^T) = S_N^{-1}+\phi(x_{N+1})\phi(x_{N+1})^T$$

Substituting this into $\sigma^2_{N+1}(x)$ and using the above matrix identity we get:

$$\sigma^2_{N+1}(x) = \frac{1}{\beta} + \phi(x)^TS_{N+1}\phi(x)$$

$$= \frac{1}{\beta} + \phi(x)^T(S_N^{-1} + \phi(x_{N+1})\phi(x_{N+1})^T)^{-1}\phi(x)$$

$$= \frac{1}{\beta} + \phi(x)^T(S_N - \frac{S_N\phi(x_{N+1})\phi(x_{N+1})^TS_N}{1 + \phi(x_{N+1})^TS_N\phi(x_{N+1})})\phi(x)$$

$$= \frac{1}{\beta} + \phi(x)^TS_N\phi(x) - \frac{\phi(x)^TS_N\phi(x_{N+1})\phi(x_{N+1})^TS_N\phi(x)}{1 + \phi(x_{N+1})^TS_N\phi(x_{N+1})} \tag{7}$$

The nominator of the last fraction is greather than zero. Now note that the S_N^{-1} is positive semi-definite since β is greater than zero, α is greater than zero, and:

$$x^T S_N^{-1} x = x^T(\alpha I + \beta \Phi_N^T \Phi_N)x = x^T \alpha I x + x^T \beta \Phi_N^T \Phi_N x = \alpha x^T I x + \beta(\Phi_N x)^T \Phi_N x$$

A matrix A has an eigenvalue if and only if A^{-1} has eigenvalue $\frac{1}{\lambda}$, as

$$Av = \lambda v \Leftrightarrow A^{-1}Av = \lambda A^{-1}v \Leftrightarrow \frac{1}{\lambda}v = A^{-1}v$$

With that and the fact that the eigenvalues of positive semi-definite matrices are non-negative we have shown that S_N also is positive semi-definite. Therefore

$$\phi(x_{N+1})^T S_N \phi(x_{N+1}) \geq 0$$

and so the denominator of the last fraction in eq. (7) is positive. We thus have proven the inequality.

Problem 3

Since we spent no money, the model boils down to $y = 10 + \mathcal{N}(0, 4)$. The Gaussian is centered on zero and, therefore, we can say that the probability to get more than 10 likes is 0.5.

Problem 4

Spending 1 EUR on advertisements our model gives us

$$y = 15 + \mathcal{N}(0, 4) \quad \Rightarrow \quad \mathrm{E}[y] = 15 \tag{8}$$

with $\quad \mathrm{E}[\mathcal{N}(0, 4)] = 0$. Hence, we expect 15 likes.