Homework 4

Due 5/5/18 by midnight

E-mail:

Problem 1

For the Bayesian linear regression model show that the posterior distribution over the parameters $\tilde{\mathbf{w}}$ is given by

$$Pr(\tilde{\mathbf{w}}|\tilde{\mathbf{X}}, \bar{\mathbf{y}}) = Norm_{\tilde{\mathbf{w}}}[\sigma^{-2}\mathbf{A}^{\mathbf{-1}}\tilde{\mathbf{X}}\bar{\mathbf{y}}, \mathbf{A}^{\mathbf{-1}}]$$

where

$$\mathbf{A} = \sigma^{-2}\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T + \sigma_p^{-2}\mathbf{I}$$

Problem 2

For the Bayesian linear regression model show that the predictive distribution for a new data example $\bar{\mathbf{x}}^*$ is given by

$$Pr(y^*|\bar{\mathbf{x}}^*, \tilde{\mathbf{X}}, \bar{\mathbf{y}}) = Norm_{y^*}[\sigma^{-2}\bar{\mathbf{x}}^{*T}\mathbf{A}^{-1}\tilde{\mathbf{X}}\bar{\mathbf{y}}, \bar{\mathbf{x}}^{*T}\mathbf{A}^{-1}\bar{\mathbf{x}}^* + \sigma^2]$$

Problem 3

Show that the gradient of the log posterior probability for the logistic regression model

$$L = \sum_{i=1}^{P} y_i log(sig(a_i)) + \sum_{i=1}^{P} (1 - y_i) log(1 - sig(a_i))$$

with respect to the parameters $\tilde{\mathbf{w}}$ is given by

$$\nabla_{\tilde{\mathbf{w}}} L = - \sum_{i=1}^{P} (sig(a_i) - y_i)\bar{\mathbf{x}}_i$$

where $a_i = \bar{\mathbf{x}}_i^T \tilde{\mathbf{w}}$ and $sig()$ denotes the sigmoid function.

Problem 4

Show that the Hessian for the log likelihood of the logistic regression model is given by

$$\nabla^2_{\tilde{\mathbf{w}}} L = - \sum_{i=1}^{P} (sig(a_i))(1 - sig(a_i))\bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T$$

Problem 5 (for graduate students only)

Consider a Bayesian linear regression model, and suppose that we have already observed P data points, so the posterior distribution over $\tilde{\mathbf{w}}$ is given as in Problem 1 by:

$$Pr(\tilde{\mathbf{w}}|\tilde{\mathbf{X}}_P, \bar{\mathbf{y}}_P) = Norm_{\tilde{\mathbf{w}}_P}[\sigma^{-2}\mathbf{A}_P^{-1}\tilde{\mathbf{X}}_P \bar{\mathbf{y}}_P, \mathbf{A}_P^{-1}] \tag{1}$$

This posterior can be regarded as the prior for the next observation. By considering an additional data point $(\bar{\mathbf{x}}_{P+1}, y_{P+1})$, and by completing the square in the exponential, show that the resulting posterior distribution is again eq.(1) but replacing P by $P+1$, where

$$\tilde{\mathbf{X}}_{P+1} = [\tilde{\mathbf{X}}_P, \bar{\mathbf{x}}_{P+1}] \quad \text{and} \quad \bar{\mathbf{y}}_{P+1} = \begin{pmatrix} \bar{\mathbf{y}}_P \\ \\ y_{P+1} \end{pmatrix}$$

References

[1] S. J. D. Prince, Computer Vision: Models Learning and Inference, Cambridge University Press, 2012.