1   Linear Algebra II

1. (20 points) Compute (by hand) the eigenvalues and the eigenvectors of the following matrix:

$$A = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

2. Given the three vectors $v_1 = (2, 0, -1)$, $v_2 = (0, -1, 0)$ and $v_3 = (2, 0, 4)$ in $\mathbb{R}^3$.

   - (10 points) Show that they form an orthogonal set under the standard Euclidean inner product for $\mathbb{R}^3$ but not an orthonormal set.

   - (10 points) Turn them into a set of vectors that will form an orthonormal set of vectors under the standard Euclidean inner product for $\mathbb{R}^3$.

3. (10 points) Suppose that A is an $n \times m$ matrix with linearly independent columns. Show that $A^T A$ is an invertible matrix.

4. (10 points) Suppose that A is an $n \times m$ matrix with linearly independent columns. Let $\bar{x}$ be a least squares solution to the system of equations $Ax = b$ (the solution of $\min_x \|Ax - b\|_2^2$). Show that $\bar{x}$ is the **unique** solution to the associated normal system $A^T A \bar{x} = A^T b$.

2   Linear Regression I

Questions in the textbook Pattern Recognition and Machine Learning:

1. (10 points) Page 174, Question 3.2

Show that the matrix
$$\mathbf{\Phi}(\mathbf{\Phi}^T\mathbf{\Phi})^{-1}\mathbf{\Phi}^T$$

takes any vector v and projects it onto the space spanned by the columns of $\mathbf{\Phi}$. Use this result to show that the least-squares solution (3.15):

$$\mathbf{w}_{\text{ML}} = (\mathbf{\Phi}^T\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\mathbf{t} \tag{3.15}$$

corresponds to an orthogonal projection of the vector $\mathbf{t}$ onto the manifold $\mathcal{S}$ as shown in Figure 3.2.

---

2. (10 points) Page 175, Question 3.7

By using the technique of completing the square, verify the result (3.49) for the posterior distribution of the parameters $\mathbf{w}$ in the linear basis function model in which $\mathbf{m}_N$ and $\mathbf{S}_N$ are defined by (3.50) and (3.51) respectively.

$$p(\mathbf{w}|\mathbf{t}) = \mathcal{N}(\mathbf{w}|\mathbf{m}_N, \mathbf{S}_N) \tag{3.49}$$

$$\mathbf{m}_N = \mathbf{S}_N(\mathbf{S}_0^{-1}\mathbf{m}_0 + \beta\boldsymbol{\Phi}^T\mathbf{t}) \tag{3.50}$$

$$\mathbf{S}_N^{-1} = \mathbf{S}_0^{-1} + \beta\boldsymbol{\Phi}^T\boldsymbol{\Phi} \tag{3.51}$$

Hint: From Bayes' theorem we have

$$p(\mathbf{w}|\mathbf{t}) \propto p(\mathbf{t}|\mathbf{w})p(\mathbf{w})$$

where the factors on the r.h.s are given by (3.10) and (3.48), respectively.

$$p(\mathbf{t}|\mathbf{X}, \mathbf{w}, \beta) = \prod_{n=1}^{N} \mathcal{N}(t_n|\mathbf{w}^T\boldsymbol{\phi}(\mathbf{x}_n), \beta^{-1}) \tag{3.10}$$

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{w}|\mathbf{m}_0, \mathbf{S}_0) \tag{3.48}$$

then plugin:

$$p(\mathbf{w}|\mathbf{t}) \propto \left[\prod_{n=1}^{N} \mathcal{N}(t_n|\mathbf{w}^T\boldsymbol{\phi}(\mathbf{x}_n), \beta^{-1})\right]\mathcal{N}(\mathbf{w}|\mathbf{m}_0, \mathbf{S}_0)$$

$$\propto \exp\left(-\frac{\beta}{2}(\mathbf{t} - \boldsymbol{\Phi}\mathbf{w})^T(\mathbf{t} - \boldsymbol{\Phi}\mathbf{w})\right)\exp\left(-\frac{1}{2}(\mathbf{w} - \mathbf{m}_0)^T\mathbf{S}_0^{-1}(\mathbf{w} - \mathbf{m}_0)\right)$$

Then expand the terms and reorganize to get to a term that is propotional to (3.49).

3. (10 points) Page 175, Question 3.10

By making use of the result (2.115) to evaluate the integral in (3.57), verify that the predictive distribution for the Bayesian linear regression model is given by (3.58) in which the input-dependent variance is given by (3.59).

$$p(t|\mathbf{t}, \alpha, \beta) = \int p(t|\mathbf{w}, \beta)p(\mathbf{w}|\mathbf{t}, \alpha, \beta)d\mathbf{w} \tag{3.57}$$

$$p(t|\mathbf{x}, \mathbf{t}, \alpha, \beta) = \mathcal{N}(t|\mathbf{m}_N^T\boldsymbol{\phi}(\mathbf{x}), \sigma_N^2(\mathbf{x})) \tag{3.58}$$

$$\sigma_N^2(\mathbf{x}) = \frac{1}{\beta} + \boldsymbol{\phi}(\mathbf{x})^T\mathbf{S}_N\boldsymbol{\phi}(\mathbf{x}). \tag{3.59}$$

Marginal and Conditional Gaussians Given a marginal Gaussian distribution for $\mathbf{x}$ and a conditional Gaussian distri- bution for y given $\mathbf{x}$ in the form

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1}) \tag{2.113}$$

$$p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}|\mathbf{A}\mathbf{x} + \mathbf{b}, \mathbf{L}^{-1}) \tag{2.114}$$

the marginal distribution of y and the conditional distribution of x given y are given by

$$p(\mathbf{y}) = \mathcal{N}(\mathbf{y}|\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{L}^{-1} + \mathbf{A}\boldsymbol{\Lambda}^{-1}\mathbf{A}^T) \tag{2.115}$$

$$p(\mathbf{x}|\mathbf{y}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\Sigma}\{\mathbf{A}^T\mathbf{L}(\mathbf{y} - \mathbf{b}) + \boldsymbol{\Lambda}\boldsymbol{\mu}\}, \boldsymbol{\Sigma}) \tag{2.116}$$

where

$$\boldsymbol{\Sigma} = (\boldsymbol{\Lambda} + \mathbf{A}^T\mathbf{L}\mathbf{A})^{-1} \tag{2.117}$$

Hint: Using (3.3), (3.8) and (3.49),

$$y(\mathbf{x}, \mathbf{w}) = \mathbf{w}^T\boldsymbol{\phi}(\mathbf{x}) \tag{3.3}$$

$$p(t|\mathbf{x}, \mathbf{w}, \beta) = \mathcal{N}(t|y(\mathbf{x}, \mathbf{w}), \beta^{-1}) \tag{3.8}$$

$$p(\mathbf{w}|\mathbf{t}) = \mathcal{N}(\mathbf{w}|\mathbf{m}_N, \mathbf{S}_N) \tag{3.49}$$

we can re-write (3.57) as

$$p(t|\mathbf{x}, \mathbf{t}, \alpha, \beta) = \int p(t|\mathbf{x}, \mathbf{w}, \beta)p(\mathbf{w}|\mathbf{t}, \alpha, \beta)d\mathbf{w} = \int \mathcal{N}(t|\boldsymbol{\phi}(\mathbf{x})^T\mathbf{w}, \beta^{-1})\mathcal{N}(\mathbf{w}|\mathbf{m}_N, \mathbf{S}_N)d\mathbf{w}$$

By matching the first factor of the integrand with (2.114) and the second factor with (2.113), we obtain the desired result directly from (2.115).

4. (10 points) Page 175, Question 3.11

We have seen that, as the size of a data set increases, the uncertainty associated with the posterior distribution over model parameters decreases. Make use of the matrix identity (Appendix C)

$$(\mathbf{M} + \mathbf{v}\mathbf{v}^T) = \mathbf{M}^{-1} - \frac{(\mathbf{M}^{-1}\mathbf{v})(\mathbf{v}^T\mathbf{M}^{-1})}{1 + \mathbf{v}^T\mathbf{M}^{-1}\mathbf{v}} \tag{3.110}$$

to show that the uncertainty $\sigma_N^2(\mathbf{x})$ associated with the linear regression function given by (3.59) satisfies

$$\sigma_{N+1}^2(\mathbf{x}) \leq \sigma_N^2(\mathbf{x})$$

Hint: From (3.59)

$$\sigma_{N+1}^2(\mathbf{x}) = \frac{1}{\beta} + \boldsymbol{\phi}(\mathbf{x})^T\mathbf{S}_{N+1}\boldsymbol{\phi}(\mathbf{x}) \tag{3.59}$$

where $\mathbf{S}_{N+1}$ is given by (r.f. Question 3.8):

$$\mathbf{S}_{N+1}^{-1} = \mathbf{S}_N^{-1} + \beta\boldsymbol{\phi}_{N+1}\boldsymbol{\phi}_{N+1}^T. \tag{P}$$

From (P) and (3.110) we get a relationship between $\mathbf{S}_{N+1}$ and $\mathbf{S}_N$, and then we can plugin to (3.59), and finally get a relationship:

$$\sigma_{N+1}^2(\mathbf{x}) = \frac{1}{\beta} + \boldsymbol{\phi}(\mathbf{x})^T \mathbf{S}_{N+1} \boldsymbol{\phi}(\mathbf{x})$$
$$= \sigma_N^2(\mathbf{x}) + \square$$

Analyze $\square$ to complete the proof.