< Linear Model for Regression>

HW9# Exercise 3.1, 3.2, 3.3, 3.6, 3.7, 3.8

3.1) show that 'tanh' function and logistic sigmoid function (3.6)

$$y(x, w) = w_0 + \sum_{j=1}^{M} w_j \sigma\left(\frac{x-\mu_j}{s}\right) \quad \cdots (3.101) \qquad \tanh(a) = 2\sigma(2a) - 1 \quad (3.100)$$

equipvalent to a linear combination 'tanh' function

$$y(x, u) = u_0 + \sum_{j=1}^{M} u_j \tanh\left(\frac{x-\mu_j}{s}\right)$$

new parameters $\{u_1, u_2, \cdots, u_M\}$ original parameters $\{w_1, \cdots, w_M\}$

sol)
$$2\sigma(2a) - 1 = \frac{2}{1+e^{-2a}} - 1 = \frac{2}{1+e^{-2a}} - \frac{1+e^{-2a}}{1+e^{-2a}} = \frac{1-e^{-2a}}{1+e^{-2a}} = \left(\frac{1-e^{-2a}}{1+e^{-2a}}\right)\cdot\frac{e^a}{e^a}$$

$a_j = (x-\mu_j)/2s \quad \cdots (3.101)$

$$= \frac{e^a - e^{-a}}{e^a + e^{-a}} = \tanh(a)$$

$$y(x, w) = w_0 + \sum_{j=1}^{M} w_j \sigma(2a_j)$$

$$= w_0 + \sum_{j=1}^{M} \frac{w_j}{2}(2\sigma(2a_j) - 1 + 1) = u_0 + \sum_{j=1}^{M} u_j \tanh(a_j)$$

where $u_j = w_j/2$, $u_0 = w_0 + \sum_{j=1}^{M} w_j/2$

$$\frac{d}{da}\tanh(a) = 1 - \tanh^2(a)$$

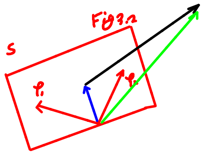
$$\left(\tanh(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}}\right)' = \frac{(e^a+e^{-a})(e^a+e^{-a}) - (e^a-e^{-a})(e^a-e^{-a})}{(e^a+e^{-a})^2} = 1 - \frac{(e^a-e^{-a})^2}{(e^a+e^{-a})^2} = 1 - \left(\frac{e^a-e^{-a}}{e^a+e^{-a}}\right)^2$$

$$= 1 - \tanh^2(a)$$

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

takes any vector v and projects it into space spanned by the columns of Φ. use this result to show that the least-squares sol (3.15). to an orthogonal projection of the vector t onto the manifold S as show Fig 3.2

Sol). $\Phi(\Phi^T\Phi)^{-1}\Phi^T v = \Phi\tilde{v} = \varphi_1\tilde{v}^{(1)} + \varphi_2\tilde{v}^{(2)} + \cdots \varphi_M\tilde{v}^{(M)}$



φ_m m-th column of Φ, $\tilde{v} = (\Phi^T\Phi)^{-1}\Phi^T v$ By comparing.

$$y = \Phi w_{ML} = \Phi(\Phi^T\Phi)^{-1}\Phi^T t$$

to projection of t onto the space spanned by the columns of Φ.

onto subspace spanned by basis fun $\phi_j(x)$

least-squares orthogonal projection.

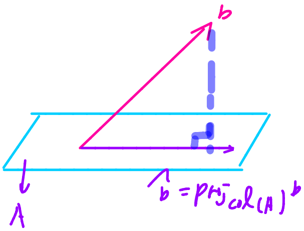
$$\Phi(\Phi^T\Phi)^{-1}\Phi^T\varphi_j = [\Phi(\Phi^T\Phi)^{-1}\Phi^T\Phi]_j = \varphi_j$$

$$\therefore (y-t)^T\varphi_j = (\Phi w_{ML} - t)^T\varphi_j = t^T(\Phi(\Phi^T\Phi)^{-1}\Phi^T - I)^T\varphi_j = 0$$

$(y-t)$, ortogonal to every column of Φ and hence is Orthogonal to S.

least-squares solution.

Find best approximation. $\hat{b} = \mathrm{proj}_{\mathrm{col}A}\, b$



$$A\hat{x} = \hat{b}$$

3.3) consider data set in which each data point t_n is associated with a weighting factor $r_n > 0$. so that the sum of squares error function becomes.

$$E_D(w) = \frac{1}{2} \sum_{n=1}^{N} r_n \{t_n - w^T \phi(x_n)\}^2$$

w^* that minimizes this error function. I) data independent noise variance ii) replicated data points.

sol) $R = \text{diagonal}(r_1, r_2, \cdots, r_n)$ diagonal containing weighting coefficients.

weighted sum of squares cost function.

↓

$$E_D(w) = \frac{1}{2} (t - \Phi w)^T R (t - \Phi w)$$

derivative with respect to to zero, re-arranging.

$$w^* = (\Phi^T R \Phi)^{-1} \Phi^T R t$$

- loss function. → Gaussian noise model...

$R = I$.

r_n can be regarded as a precision (inverse variance)

$$E_D(w) = \frac{1}{2} \sum_{n=1}^{N} (t_n - w^T \phi(x_n))^2$$

alternatively, r_n can be regarded as an effective number of replicated observations of data point (x_n, t_n)

r_n positive integer value. $r_n > 0$.

3.6) consider a linear basis function regression model for multivariate.

target variable t having a Gaussian distribution of the form

$$p(t|W, \Sigma) = \mathcal{N}(t|y(x, W), \Sigma) \quad (3.107)$$

$$y(x, w) = W^T \phi(x) \quad (3.108)$$

input basis vector $\phi(x_n)$, target vectors t_n, $n = 1, \dots, N$

Show that likelihood sol W_{ML} parameter matrix W (3.15)

isotropic noise distribution. N.t independent of convirience matrix Σ

show that Maximum likehood sol Σ. is given by.

$$\Sigma = \frac{1}{N} \sum_{n=1}^{N} (t_n - W_{ML}^T \phi(x_n)) (t_n - W_{ML}^T \phi(x_n))^T \quad (3.109)$$

sol) likelihood function,

$$\ln L(W, \Sigma) = -\frac{N}{2} \ln|\Sigma| - \frac{1}{2} \sum_{n=1}^{N} (t_n - W^T \phi(x_n))^T \Sigma^{-1} (t_n - W^T \phi(x_n))$$

add we set derivative respect to w equal to zero,

$$0 = -\sum_{n=1}^{N} \Sigma^{-1} (t_n - W^T \phi(x_n)) \phi(x_n)^T$$

$$\Phi^T \Phi W = \Phi^T T$$

$$W_{ML} = (\Phi^T \Phi)^{-1} \Phi^T t \quad (3.15)$$

normal equation

least-squares problem.

Φ

$$\Phi = \begin{pmatrix} \phi_0(x_1) & \cdots & \phi_{M-1}(x_1) \\ \vdots & & \vdots \\ \phi_0(x_N) & \cdots & \phi_{M-1}(x_N) \end{pmatrix}$$

$N \times M$ matrix

ML sol for Σ

$$\Sigma = \frac{1}{N} \sum_{n=1}^{N} (t_n - W_{ML}^T \phi(x_n)) (t_n - W_{ML}^T \phi(x_n))^T$$

1. posterior : Given object, objection of the probability distribution.

2. likelihood : Given object don't know, know suppose. "

3. pirior : experience.

3.7) By using Technique of completing the square, verify the resault (3.49) for posterior distribution of the parameter w in the linear basis function model in which m_N and S_N (3.50), (3.51)

$$(3.50) \quad m_N = S_N(S_0^{-1}m_0 + \beta\Phi^T t) \qquad (3.51) \quad S_N^{-1} = S_0^{-1} + \beta\Phi^T\Phi$$

sol) Bayes' theorem, $p(w|t) \propto p(t|w)\,p(w)$.

m_0 (mean) (Variance) S_0

$$(3.48) \quad p(w) = \mathcal{N}(w|m_0, S_0)$$

$$(3.10) \quad p(t|X, w, \beta) = \prod_{n=1}^{N} \mathcal{N}(t_n|w^T\phi(x_n), \beta^{-1})$$

$$p(w|t) \propto \left[\prod_{n=1}^{N} \mathcal{N}(t_n|w^T\phi(x_n), \beta^{-1})\right]\mathcal{N}(w|m_0, S_0)$$

$$\propto \exp\left(-\frac{\beta}{2}(t-\Phi w)^T(t-\Phi w)\right)$$

$$\exp\left(-\frac{1}{2}(w-m_0)^T S_0^{-1}(w-m_0)\right) = \exp\Big(-\frac{1}{2}\big(w^T(S_0^{-1}+\beta\Phi^T\Phi)w - \beta t^T\Phi w$$

$$-\beta w^T\Phi^T t + \beta t^T t\, m_0^T S_0^{-1} w - w^T S_0^{-1} m_0 + m_0^T S_0^{-1} m_0\big)\Big)$$

$$= \exp\left(-\frac{1}{2}\left(w^T(S_0^{-1}+\beta\Phi^T\Phi)w - (S_0^{-1}m_0+\beta\Phi^T t)^T w - w^T(S_0^{-1}m_0+\beta\Phi^T t) + \beta t^T t + m_0^T S_0^{-1} m_0\right)\right)$$

$$= \exp\left(-\frac{1}{2}(w-m_N)^T S_N^{-1}(w-m_N)\right)\exp\left(-\frac{1}{2}\left(\beta t^T t + m_0^T S_0^{-1} m_0 - m_N^T S_N^{-1} m_N\right)\right)$$

Gaussian distribution over w

Normalization factor.

7.81 consider the linear basis function model Section 3.1 suppose that we have already observed N data points, so that the posterior distribution over w (3.49) This posterior can be regared. By considering an additional data point (x_{N+1}, t_{N+1}) S_N replaced by S_{N+1}, m_N replaced by m_{N+1}

sol) $p(w) = N(w|m_N, S_N)$ likelihood $p(t_{N+1}|x_{N+1}, w) = \left(\frac{\beta}{2\pi}\right)^{1/2} \exp\left(-\frac{\beta}{2}(t_{N+1} - w^T\phi_{N+1})^2\right)$

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

$$p(w|t_{N+1}, x_{N+1}, m_N, S_N) \propto \exp\left(-\frac{1}{2}(w-m_N)^T S_N^{-1}(w-m_N) - \frac{1}{2}\beta(t_{N+1} - w^T\phi_{N+1})^2\right)$$

we can expand argument. exponential $-1/2$ factors

$$(w-m_N)^T S_N^{-1}(w-m_N) + \beta(t_{N+1} - w^T\phi_{N+1})^2$$

$$= w^T S_N^{-1} w - 2w^T S_N^{-1} m_N + \beta w^T \phi_{N+1}^T \phi_{N+1} w - 2\beta w^T \phi_{N+1} t_{N+1} + \text{const}$$

$$= w^T(S_N^{-1} + \beta\phi_{N+1}\phi_{N+1}^T) w - 2w^T(S_N^{-1}m_N + \beta\phi_{N+1}t_{N+1}) + \text{const}$$

Cost denotes

$$p(w|t_{N+1}, x_{N+1}, m_N, S_N) = N(w|m_{N+1}, S_{N+1})$$

$$S_{N+1}^{-1} = S_N^{-1} + \beta\phi_{N+1}\phi_{N+1}^T, \qquad m_{N+1} = S_{N+1}(S_N^{-1}m_N + \beta\phi_N t_{N+1})$$