PAUL EMERICK NGOUCHET (1)

HW3 Machine Learning CS 542

1. Written problems.

(a) Bishop 5.3.

$$p(T|X, w, \Sigma) = \prod_{n=1}^{N} \mathcal{N}(t_n | y(x_n, w), \Sigma)$$

$$\ln p(T|X, w, \Sigma)$$

$$= -\frac{N}{2}(\ln|\Sigma| + k\ln(2\pi)) - \frac{1}{2}\sum_{n=1}^{N}(t_n - y_n)^T \Sigma^{-1}(t_n - y_n)$$

where $y_n = y(x_n, w)$.

$$E(w) = \frac{1}{2}\sum_{n=1}^{N}(t_n - y_n)^T \Sigma^{-1}(t_n - y_n)$$

$$-\frac{N}{2}\ln|\Sigma| - \frac{1}{2}\sum_{n=1}^{N}(t_n - y_n)^T \Sigma^{-1}(t_n - y_n)$$

$$\Rightarrow -\frac{N}{2}\ln|\Sigma| - \frac{1}{2}\mathrm{Tr}\left[\Sigma^{-1}\sum_{n=1}^{N}(t_n - y_n)(t_n - y_n)^T\right]$$

We can maximize this by setting the derivative w.r.t Σ^{-1} to zero $\Rightarrow$ $\boxed{\Sigma = \frac{1}{N}\sum_{n=1}^{N}(t_n - y_n)(t_n - y_n)^T}$

(2)

Thus the optimal value for Σ depends on w through y_n.
To address the dependence between w and Σ, we could try an iterative scheme, alternating between updates of W and Σ until some convergence criterion is reached

(b) <u>Bishop 5.4.</u>

$t \in \{0, 1\}$ data set label
$k \in \{0, 1\}$ true class label ; $y(x, w) = P(k=1|X)$

$$P(t=1|x) = \sum_{k=0}^{1} P(t=1|k)\, P(k|x) = (1-\varepsilon)\, y(x,w) + \varepsilon(1-y(x,w)).$$

$$P(t|X) = P(t=1|X)^t \,(1-P(t=1|X))^{1-t}.$$

$$\boxed{E(w) = -\sum_{n=1}^{N} \{ t_n \ln[(1-\varepsilon)\, y(x_n, w) + \varepsilon(1-y(x_n,w))] + (1-t_n) \ln[1-(1-\varepsilon)\, y(x_n,w) - \varepsilon(1-y(x_n,w))]\}}$$

(3)

5.26.

5.201 $\quad \Omega_n = \frac{1}{2}\sum_k (\mathcal{G} y_k)^2 \Big|_{x_n}$

5.202 → 5.201 and 5.70

$\cancel{\Omega}$

$$\Rightarrow \Omega_n = \frac{1}{2}\sum_k \Big(\sum_i \tau_{ni} \frac{\partial y_{nk}}{\partial x_{ni}}\Big)^2 \qquad (188)$$

$$= \frac{1}{2}\sum_k \Big(\sum_i \tau_{ni} J_{nki}\Big) \qquad (189)$$

5.202, 5.203, 5.205

$$\beta_{nj} = \sum_i w_{ji}\,\alpha_{ni}$$

$$= \sum_i w_{ji}\,\mathcal{G} x_{ni}$$

$$= \sum_i w_{ji} \sum_{i'} \tau_{ni'} \frac{\partial x_{ni}}{\partial x_{ni'}}$$

$$= \sum_i w_{ji}\,\tau_{ni}.$$

$$\Omega_n = \frac{1}{2}\sum_k (\mathcal{G} y_{nk})^2 = \frac{1}{2}\sum_k \alpha_{nk}^2$$

$$\frac{\partial \Omega_n}{\partial w_{rs}} = \sum_k (\mathcal{G} y_{nk})\,\mathcal{G}(\delta_{nkr} z_{ns})$$

$$= \sum_k \alpha_{nk}(\phi_{nkr} z_{ns} + \delta_{nkr}\alpha_{ns})$$

(4)

(5.74)

$$\delta_{nkr} = h'(a_{nr}) \sum_l w_{lr} \delta_{nkl}$$

5.205 and 5.207.

$$\phi_{nkr} = \mathcal{G}\delta_{nkr}$$
$$= \mathcal{G}\left(h'(a_{nr}) \sum_l w_{lr} \delta_{nkl}\right)$$
$$= h''(a_{nr}) \beta_{nr} \sum_l w_{lr} \delta_{nkl} + h'(a_{nr}) \sum_l w_{lr} \phi_{nkl}$$