Shimui Ye

PS3

CS542

Cover

To achieve the best result, I decide to use hand to write down the written assignment part. (Typing in math symbols and equations in LaTex really interrupts my logic)

Shirui Ye
CS542
PS3

5.3 $p(t|X,w,\Sigma) = \prod_{n=1}^{N} \mathcal{N}(t_n|y(x_n,w),\Sigma)$ (Used Gaussian law $t_n|y(x_n,w)$ is μ, Σ is 6^2)

To mak $\prod$ into Σ, take the $\ln$

$\ln[p(t|X,w,\Sigma) = -\frac{1}{2}\sum_{n=1}^{N}(t_n-y_n)^T\Sigma^{-1}(t_n-y_n) - \frac{N}{2}(\ln|\Sigma| + \underbrace{k\ln(2\pi)}_{})$

constant we can ignore
it won't affect maximization.

we used Gaussian and in the law $y_n = y(x_n,w)$,

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

trace of matrix, sum of diagnol

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

To maximize, take derivative of Σ^{-1}

The maximum likelihood is $\Sigma = \frac{1}{N}\sum_{n=1}^{N}(t_n-y_n)(t_n-y_n)^T$

Shimi Ye

CS542

PS3

5.4 target values are $t \in \{0,1\}$.

set $t_t \in \{0,1\}$ to be true class label.

According to question, we have a network output $y(x, w)$ that represents $p(t=1|x)$

we can get below:

$$p(t=1|X) = \epsilon(1-y(x,w)) + (1-\epsilon)y(x,w)$$

Then we can get conditional prob $p(t|x)$ to be below:

$$p(t|X) = (1-p(t=1|x))^{1-t} p(t=1|x)^t$$

Finally, the error function corresponding to the negative log likelihood is:

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

Thus, the overall likelihood to the error function is obtained when $\epsilon=0$.

Shirui Ye
CS542
PS3

5.26 We want to show that the regularization term Ω can be written as a sum over patterns of terms of form

$$\Omega_n = \frac{1}{2}\sum_K (\mathcal{G} y_K)^2 \qquad \Omega_n = \frac{1}{2}\sum_K \left(\sum_i \tau_{ni} \frac{\partial y_{nk}}{\partial x_{ni}}\right)^2$$

$\sum_i \tau_i \frac{\partial}{\partial x_i} \Rightarrow$ put this in

use Jacobian Matrix $J_{ki} = \frac{\partial y_K}{\partial x_i}$

$$\Omega_n = \frac{1}{2}\sum_K \Big(\sum_i \tau_{ni} \underbrace{J_{nki}}_{J_{ki} \text{ at } x_n}\Big)^2$$

Next by acting forward propagation equations with $\mathcal{G}$, we want to show Ω_n can be evaluated by using equations 5.204

Applying $\mathcal{G} \equiv \sum_i \tau_i \frac{\partial}{\partial x_i}$ in $z_j = h(a_j)$, $a_j = \sum_i w_{ji} z_i$, and we have $\alpha_j \equiv \mathcal{G} z_j$, $\beta_j \equiv \mathcal{G} a_j$, we can get $\alpha_j = h'(a_j)\beta_j$, $\beta_j = \sum_i w_{ji} \alpha_i$, now we see B_{nl} can write in α_{ni}.

According to these 2, we can get $B_{nj} = \sum_i w_{ji}\alpha_{ni} = \sum_i w_{ji} \sum_{i'} \tau_{ni'} \frac{\partial x_{ni}}{\partial x_{ni'}} = \sum_i w_{ji} \tau_{ni}$

This is propagated forward by equation $\alpha_j = h'(a_j)\beta_j$, $\beta_j = \sum_i w_{ji} \alpha_i$

B_{nl} output layer so that Ω_n is gotten by

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

Finally, we want to show that the derivatives of Ω_n with respect to a weight w_{rs} can be written in the form $\frac{\partial \Omega_n}{\partial w_{rs}} = \sum_K \alpha_K \{\phi_{Kr} z_s + \delta_{Kr} \alpha_s\}$.

To do so, we use chain rule, $\Omega_n = \frac{1}{2}\sum_K (\mathcal{G} y_K)^2$, $\alpha_j \equiv \mathcal{G} z_j$, $\beta_j \equiv \mathcal{G} a_j$ and $\delta_{Kr} \equiv \frac{\partial y_K}{\partial a_r}$, $\phi_{Kr} \equiv \mathcal{G}\delta_{Kr}$, $\frac{\partial a_j}{\partial w_{ji}} = z_i$

We can get $\frac{\partial \Omega_n}{\partial w_{rs}} = \sum_K \alpha_{nK}(\phi_{nKr} z_{ns} + \delta_{nKr}\alpha_{ns})$

According to $\frac{\partial y_K}{\partial a_j} = h'(a_j)\sum_l w_{lj} \frac{\partial y_K}{\partial a_l}$, we know $\delta_{nKr} = h'(a_{nr})\sum_l w_{lr}\delta_{nKl}$

Now, we can get backpropagation of ϕ_{nKr}, by using above with $\alpha_j \equiv \mathcal{G} z_j$, $\beta_j = \mathcal{G} a_j$, $\delta_{Kr} \equiv \frac{\partial y_K}{\partial a_r}$, $\phi_{Kr} \equiv \mathcal{G}\delta_{Kr}$

$$\phi_{nKr} = \mathcal{G}\delta_{nKr} = h'(a_{nr})\sum_l w_{lr}\phi_{nKl} + h''(a_{nr})\beta_{nr}\sum_l w_{lr}\delta_{nKl}$$