Xiao Zhou zhouxiao@bu.edu
CS 542 Problem Set 3
Discussed with classmates
3/27/2017

1. Written Problems
   a. Bishop 5.3
   $p(T|X,w,\Sigma)=\prod_{n=1}^{N}N(t_n|y(x_n,w),\Sigma)$
   Let $y_n=y(x_n,w)$ and $K=dimensionality$ of y and $t, then$
   $\ln p(T|X,w,\Sigma)=-\frac{N}{2}\left(\ln|\Sigma|+Kln(2\pi)\right)-\frac{1}{2}\sum_{n=1}^{N}(t_n-y_n)^T\Sigma^{-1}(t_n-y_n)$
   If Σ is fixed and known, then we would have
   $$E(w)=\frac{1}{2}\sum_{n=1}^{N}\left(t_n-y_n\right)^T\Sigma^{-1}\left(t_n-y_n\right)$$
   If otherwise, then we have
   $$\ln p(T|X,w,\Sigma)=-\frac{N}{2}\ln|\Sigma|-\frac{1}{2}\sum_{n=1}^{N}(t_n-y_n)^T\Sigma^{-1}(t_n-y_n)$$
   $$=-\frac{N}{2}\ln|\Sigma|-\frac{1}{2}Tr[\Sigma^{-1}\sum_{n=1}^{N}(t_n-y_n)(t_n-y_n)^T]$$
   By setting derivative w.r.t. $\Sigma^{-1}=0$,
   $$\Sigma=\frac{1}{N}\sum_{n=1}^{N}\left(t_n-y_n\right)\left(t_n-y_n\right)^T$$

   b. Bishop 5.4
   Let $k\in\{0,1\}$ be the true class label.
   $$p(t=1|x)=p(t=1|k=0)p(k=0|x)+p(t=1|k=1)p(k=1|x)$$
   $$=(1-\epsilon)y(x,w)+\epsilon\left(1-y(x,w)\right)$$
   $$p(t|x)=p(t=1|x)^t\left(1-p(t=1|x)\right)^{1-t}$$
   $$E(w)=-\sum_{n=1}^{N}t_n\ln\left[(1-\epsilon)y(x_n,w)+\epsilon\left(1-y(x_n,w)\right)\right]$$
   $$+(1-t_n)\ln\left[1-(1-\epsilon)y(x_n,w)-\epsilon\left(1-y(x_n,w)\right)\right]$$

   c. Bishop 5.26
   $$\Omega_n=\frac{1}{2}\sum_k(\mathcal{G}\mathrm{y_k})^2\bigg|_{x_n}$$
   Since $\mathrm{J_{ki}}\equiv\frac{\partial y_k}{\partial x_i}$, we have
   $\Omega_\mathrm{n}=\frac{1}{2}\sum_k\left(\sum_i\tau_{\mathrm{ni}}\frac{\partial y_{nk}}{\partial x_{ni}}\right)^2$, which is equivalent to (5.128)

   Now, we have $\mathrm{z_j}=h\left(a_j\right), a_j=\sum_i w_{ji}z_i$, $\alpha_\mathrm{j}=h'\left(\alpha_j\right)\beta_j, \beta_j=\sum_i w_{ji}\alpha_i$, $\alpha_\mathrm{j}\equiv\mathcal{G}\mathrm{z_j}, \beta_\mathrm{j}\equiv\mathcal{G}\mathrm{a_j}$
   $$\beta_{\mathrm{nj}}=\sum_i w_{ji}\alpha_{ni}=\sum_i w_{ji}\mathcal{G}\mathrm{x_{ni}}=\sum_i w_{ji}\sum_{i'}\tau_{\mathrm{ni'}}\frac{\partial y_{nk}}{\partial x_{ni'}}=\sum_i w_{ji}\tau_{\mathrm{ni}}$$

As τ_n is propagated forward by $\alpha_j = h'(\alpha_j)\beta_j, \beta_j = \sum_i w_{ji}\alpha_i$, we have β_{nl} for the output layer, and Ω_n can the be computed as

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

For the derivatives,

Let $\delta_{kr} \equiv \frac{\partial y_k}{\partial a_r}, \phi_{kr} \equiv \mathcal{G}\delta_{kr}$

Then $\frac{\partial \Omega_n}{\partial w_{rs}} = \sum_k (\mathcal{G}y_{nk})\mathcal{G}(\delta_{nkr} z_{ns}) = \sum_k \alpha_{nk}(\phi_{kr} z_s + \delta_{kr}\alpha_s)$

With 5.74, we can get

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

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

2. Programming

Available in ‘starter’ folder.