MFDNN

Homework 12

24/06/15

1. (Workings)

$$L(\underline{\theta},\underline{\phi}) = \frac{1}{N}\sum_{i=1}^{N}\log(1+\exp(-Y_i(\underline{X}_i-\underline{\phi})^T\underline{\theta})) - \frac{\lambda}{2}\|\underline{\phi}\|^2$$

$$= \frac{1}{N}\sum_{i=1}^{N}\log\left(1+\exp\left(-Y_i\sum_{k=1}^{p}((X_i)_k-\phi_k)\theta_k\right)\right) - \frac{\lambda}{2}\sum_{k=1}^{p}\phi_k^2$$

$$\nabla_\phi L(\underline{\theta},\underline{\phi}) = \frac{1}{N}\sum_{i=1}^{N}\frac{Y_i\underline{\theta}\exp(-Y_i(\underline{X}_i-\underline{\phi})^T\underline{\theta})}{1+\exp(-Y_i(\underline{X}_i-\underline{\phi})^T\underline{\theta})} - \lambda\underline{\phi}$$

$$\nabla_\theta L(\underline{\theta},\underline{\phi}) = \frac{1}{N}\sum_{i=1}^{N}\frac{Y_i(\underline{\phi}-\underline{X}_i)\exp(-Y_i(\underline{X}_i-\underline{\phi})^T\underline{\theta})}{1+\exp(-Y_i(\underline{X}_i-\underline{\phi})^T\underline{\theta})}$$

2. $$f(u) = \begin{cases} u\log\frac{u}{\lambda+u}+\lambda\log\frac{\lambda}{\lambda+u}+(1+\lambda)\log(1+\lambda)-\lambda\log\lambda, & u\geq 0 \\ \infty & \text{, otherwise}\end{cases}$$

$$\frac{\partial}{\partial u}\left(u(\lambda+u)^{-1}\right) = \frac{1}{\lambda+u}-\frac{u}{(\lambda+u)^2} \quad \times(\lambda+u)$$

$$\frac{\partial}{\partial u}\left(\lambda(\lambda+u)^{-1}\right) = \frac{-\lambda}{(\lambda+u)^2} \quad \times(\lambda+u)$$

$$\frac{\partial}{\partial u}(tu-f(u)) = t-\left(\log\frac{u}{\lambda+u}+\frac{\lambda}{\lambda+u}-\frac{\lambda}{\lambda+u}\right)=0 \text{ when maximised}$$

$$\Leftrightarrow e^t = \frac{u}{\lambda+u} \Leftrightarrow \lambda e^t+ue^t = u \Leftrightarrow u(1-e^t)=\lambda e^t \Leftrightarrow u=\frac{\lambda}{e^{-t}-1}$$

$$f^*(t) = \sup_{u\in\mathbb{R}}\{tu-f(u)\} \overset{\text{if } u\geq 0 \Leftrightarrow t<0}{=} t\lambda(e^{-t}-1)^{-1}-\left(\lambda(e^{-t}-1)^{-1}\log\frac{(e^{-t}-1)^{-1}}{1+(e^{-t}-1)^{-1}}+\lambda\log\frac{1}{1+(e^{-t}-1)^{-1}}+(1+\lambda)\log(1+\lambda)-\lambda\log\lambda\right)$$

$$= \frac{t\lambda}{e^{-t}-1}-\frac{\lambda}{e^{-t}-1}\log\frac{1}{e^{-t}}-\lambda\log\frac{e^{-t}-1}{e^{-t}}-(1+\lambda)\log(1+\lambda)+\lambda\log\lambda$$

$$= \frac{t\lambda}{e^{-t}-1}-\frac{t\lambda}{e^{-t}-1}-\lambda\log(1-e^t)-(1+\lambda)\log(1+\lambda)+\lambda\log\lambda$$

$$= \lambda\log\lambda-(1+\lambda)\log(1+\lambda)-\lambda\log(1-e^t) \quad \text{for } t<0\text{, otherwise } f^*(t)=\infty$$

$$\underset{\theta}{\text{minimise}}\; D_f(p_{true}\|p_\theta) \overset{\text{for infinitely expressive } D_\phi}{=} \underset{\theta}{\text{minimise}}\;\underset{\phi}{\text{maximise}}\; \mathbb{E}_{X\sim p_{true}}[\rho(D_\phi(X))]-\mathbb{E}_{\tilde{X}\sim p_\theta}[f^*(\rho(D_\phi(\tilde{X})))]\text{, by slide 131}$$

With $\rho(r)=\log(r)$ to ensure $t<0$ since $D_\phi:\mathbb{R}^n\to(0,1)$, (as on slide 134)

$$\underset{\theta}{\text{minimise}}\; D_f(p_{true}\|p_\theta) = \underset{\theta}{\text{minimise}}\;\underset{\phi}{\text{maximise}}\; \mathbb{E}_{X\sim p_{true}}[\log D_\phi(X)]+\mathbb{E}_{\tilde{X}\sim p_\theta}[\lambda\log(1-D_\phi(\tilde{X}))]-\lambda\log\lambda+(1+\lambda)\log(1+\lambda)$$

$$= \underset{\theta}{\text{minimise}}\;\underset{\phi}{\text{maximise}}\; \mathbb{E}_{X\sim p_{true}}[\log D_\phi(X)]+\lambda\mathbb{E}_{\tilde{X}\sim p_\theta}[\log(1-D_\phi(\tilde{X}))] \quad \blacksquare$$