MFDNN

Homework 2

24/03/14

1. (Workings)

Given the minimisation problem,

$$\underset{\underline{\theta} \in \mathbb{R}^p}{\text{minimise}} \ \frac{1}{N}\sum_{i=1}^{N} \log(1+\exp(-Y_i \underline{X}_i^T \underline{\theta})) \ =: F(\theta)$$

Define $f_i(\underline{\theta}) = \log(1+\exp(-Y_i \underline{X}_i^T \underline{\theta}))$ $\qquad \left(\frac{1}{N}\sum_{i=1}^{N} f_i(\underline{\theta}) = F(\underline{\theta})\right)$

$\hookrightarrow \sum_{j=1}^{p} X_{ij}^T \theta_j$

$$\frac{\partial}{\partial \theta_k}(f_i(\underline{\theta})) = -Y_i X_{ik}^T \exp(-Y_i \underline{X}_i^T \underline{\theta}) \frac{1}{1+\exp(-Y_i \underline{X}_i^T \underline{\theta})}$$

$$\Rightarrow \underline{\nabla} f_i(\underline{\theta}) = -\sum_{k=1}^{p} \frac{\exp(-Y_i \underline{X}_i^T \underline{\theta})}{1+\exp(-Y_i \underline{X}_i^T \underline{\theta})} Y_i X_{ik}^T \underline{e}_k$$

$$= -\frac{\exp(-Y_i \underline{X}_i^T \underline{\theta})}{1+\exp(-Y_i \underline{X}_i^T \underline{\theta})} Y_i \underline{X}_i^T \longrightarrow \text{code}$$

SGD involves picking an i randomly each iteration

2. (Workings)

$$f_i(\underline{\theta}) = \max\{0, 1 - Y_i \underline{X}_i^T \underline{\theta}\} + \lambda \|\underline{\theta}\|^2$$

If $1 - Y_i \underline{X}_i^T \underline{\theta} > 0$,

$$\frac{\partial}{\partial \theta_k}(f_i(\underline{\theta})) = -Y_i X_{ik}^T + 2\lambda \theta_k$$

$$\Rightarrow \underline{\nabla} f_i(\underline{\theta}) = -Y_i \underline{X}_i^T + 2\lambda \underline{\theta} \longrightarrow \text{code}$$

If $1 - Y_i \underline{X}_i^T \underline{\theta} < 0$,

$$\underline{\nabla} f_i(\underline{\theta}) = 2\lambda \underline{\theta} \longrightarrow \text{code}$$

4. First consider the set of non-negative real numbers, $\mathbb{R}_+ := \{x \in \mathbb{R} \mid x \geq 0\} = [0, \infty)$

Given $x_1, x_2 \in \mathbb{R}_+$ and $p \in (0,1)$, define $x_3 = px_1 + (1-p)x_2$

Is $x_3 \in \mathbb{R}_+$? $\quad x_3 \overset{?}{\geq} 0 \Leftrightarrow$

$$px_1 + (1-p)x_2 \overset{?}{\geq} 0$$
$$x_1 + \frac{1-p}{p}x_2 \overset{?}{\geq} 0 \qquad (p>0)$$
$$\left(\frac{1}{p}-1\right)x_2 \overset{?}{\geq} -x_1$$
$$\left(1-\frac{1}{p}\right)x_2 \overset{?}{\leq} x_1$$

$$\frac{1}{p} > 1 \ \forall p \in (0,1) \Rightarrow 1 - \frac{1}{p} < 0 \ \forall p \Rightarrow \left(1-\frac{1}{p}\right)x_2 \leq 0 \leq x_1$$

Since $x_2 \in \mathbb{R}_+$, $x_2 \geq 0$ so whole inequality must hold $\Rightarrow x_3 \in \mathbb{R}_+$ too

$\Rightarrow$ So $\mathbb{R}_+$ is convex.

Now we consider $\varphi : \mathbb{R}_+ \to \mathbb{R}$

$\varphi(x) = -\log(x)$ and try to show φ is convex.

We have already show $\mathbb{R}_+$ is convex above so we proceed with the second condition.

Given $x_1, x_2 \in \mathbb{R}_+$ and $p \in (0,1)$, we need to check the inequality:

$$-\log(px_1 + (1-p)x_2) \overset{?}{\leq} -p\log(x_1) - (1-p)\log(x_2)$$

$$\log(px_1 + (1-p)x_2) \overset{?}{\geq} p\log(x_1) + (1-p)\log(x_2) = p\log\left(\frac{x_1}{x_2}\right) + \log(x_2)$$

$$px_1 + (1-p)x_2 \overset{?}{\geq} \exp\left(p\log\left(\frac{x_1}{x_2}\right) + \log(x_2)\right)$$

since exp is strictly increasing so preserves inequality

$$= \exp\left(p\log\left(\frac{x_1}{x_2}\right)\right)\exp(\log(x_2))$$
$$= \frac{x_1^p}{x_2^p} \cdot x_2$$

$$px_1 + (1-p)x_2 \overset{?}{\geq} x_1^p \, x_2^{1-p}$$

By the weighted Arithmetic Mean - Geometric Mean inequality:

$w_1 = p$, $w_2 = 1-p$; $x_1, x_2, w_1, w_2 \geq 0$; $w = w_1 + w_2 = 1 > 0$ so

$$\frac{w_1 x_1 + w_2 x_2}{w} \geq \sqrt[w]{x_1^{w_1} x_2^{w_2}}$$ holds.

$\Rightarrow$ Our inequality is true $\Rightarrow -\log(x)$ is convex from $\mathbb{R}_+ \to \mathbb{R}$

$$D_{KL}(p \| q) = \sum_{i=1}^n p_i \log\left(\frac{p_i}{q_i}\right)$$
$$= \mathbb{E}_I\left[\log\left(\frac{p_I}{q_I}\right)\right] \quad \text{for r.v. } I \text{ s.t. } \mathbb{P}(I=i) = p_i$$
$$= \mathbb{E}_I[-\log(q_I/p_I)] \geq -\log(\mathbb{E}_I[q_I/p_I])$$

by Jensen's inequality since $-\log(x)$ is convex and $q_I/p_I \in \mathbb{R}_+$ since q, p are pmfs satisfying $q_i, p_i \geq 0$ (for 0 conventions, see slides)

$$= -\log\left(\sum_{i=1}^n p_i \cdot q_i/p_i\right) = -\log\left(\sum_{i=1}^n q_i\right)$$
$$= -\log(1) = 0 \ \blacksquare$$

since q is a pmf satisfying $\sum q_i = 1$

5. It turns out that we can actually prove $\varphi(x) := -\log(x)$ is <u>strictly</u> convex.
For the step where we used the weighted AM-GM inequality, there is equality iff $x_1 = x_2$.
Since in the first step of the proof for strict convexity we take $x_1, x_2 \in \mathbb{R}_+$ subject to $x_1 \neq x_2$, we can conclude:

$$px_1 + (1-p)x_2 > x_1^p x_2^{1-p}$$ i.e. that $-\log(x)$ is <u>strictly</u> convex from $\mathbb{R}_+ \to \mathbb{R}$

Now we consider $D_{KL}(p \| q)$ for $p, q \in \mathbb{R}^n$ probability mass functions as before but with $p \neq q$
We have,

$$D_{KL}(p\|q) = \sum_{i=1}^n p_i \log(p_i/q_i)$$

$$= \mathbb{E}_I[\log(p_I/q_I)]$$ for I a r.v. s.t. $P(I=i) =$

Observe that r.v. $X := p_I/q_I \in \mathbb{R}_+$ is non-constant if $p \neq q$

$$= \mathbb{E}_I[-\log(q_I/p_I)] > -\log(\mathbb{E}_I[q_I/p_I])$$ in the case X is indeed non-constant by Strict Jensen's inequality

$$= -\log\left(\sum_{i=1}^n p_i \cdot q_i/p_i\right)$$

$$= -\log\left(\sum_{i=1}^n q_i\right) = -\log(1)$$ since q a pmf s.t. $\sum_{i=1}^n q_i = 1$

$$= 0$$

$$\Rightarrow D_{KL}(p\|q) > 0 \text{ if } p \neq q$$ ∎

6. $f_{\underline{\theta}}(x) = f_{\underline{a},\underline{b},\underline{u}}(x) = \sum_{j=1}^{p} u_j \sigma(a_j x + b_j)$

$$\frac{\partial}{\partial u_k}\left(f_{\underline{a},\underline{b},\underline{u}}(x)\right) = \sigma(a_k x + b_k) \Rightarrow \nabla_{\underline{u}} f_{\underline{\theta}}(x) = \sum_{k=1}^{p} \sigma(a_k x + b_k)\underline{e}_k$$

$$= \underline{\sigma}(\underline{a}x + \underline{b})$$ following given notation

$$\frac{\partial}{\partial b_k}\left(f_{\underline{a},\underline{b},\underline{u}}(x)\right) = u_k \sigma'(a_k x + b_k) \Rightarrow \nabla_{\underline{b}} f_{\underline{\theta}}(x) = \sum_{k=1}^{p} u_k \sigma'(a_k x + b_k)\underline{e}_k$$

$$= \underline{u} \odot \underline{\sigma}'(\underline{a}x + \underline{b})$$

$$= \underline{\sigma}'(\underline{a}x + \underline{b}) \odot \underline{u}$$

$$= \mathrm{diag}(\underline{\sigma}'(\underline{a}x + \underline{b}))\underline{u}$$

$$\begin{pmatrix} \sigma_1' & 0 & \cdots & 0 \\ 0 & \sigma_2' & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & \sigma_p' \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_p \end{pmatrix} = \sigma_1' u_1 + \sigma_2' u_2 + \dots + \sigma_p' u_p$$

$$= \underline{\sigma}' \odot \underline{u}$$

$$\frac{\partial}{\partial a_k}\left(f_{\underline{a},\underline{b},\underline{u}}(x)\right) = u_k x \sigma'(a_k x + b_k) \Rightarrow \nabla_{\underline{a}} f_{\underline{\theta}}(x) = \sum_{k=1}^{p} u_k x_k \sigma'(a_k x + b_k)\underline{e}_k$$

$$= \sum_{k=1}^{p} \sigma'(a_k x + b_k) u_k x \underline{e}_k$$

$$= (\underline{\sigma}'(\underline{a}x + \underline{b}) \odot \underline{u}) x$$ like above

$$= \mathrm{diag}(\underline{\sigma}'(\underline{a}x + \underline{b}))\underline{u}\, x \quad \blacksquare$$

7. (Workings)

$$\ell_{\underline{\theta}}(X,Y) = \tfrac{1}{2}\left(f_{\underline{\theta}}(X) - Y\right)^2$$

$$\nabla_{\underline{\theta}} \ell_{\underline{\theta}}(x,y) = \nabla_{\underline{\theta}} f_{\underline{\theta}}(X)\left(f_{\underline{\theta}}(X) - Y\right)$$

$$= \left(f_{\underline{\theta}}(X) - Y\right) \underbrace{\nabla_{\underline{\theta}} f_{\underline{\theta}}(X)}_{\{\nabla_{\underline{a}} f_{\underline{\theta}},\, \nabla_{\underline{b}} f_{\underline{\theta}},\, \nabla_{\underline{u}} f_{\underline{\theta}}\}}$$ i.e. the three vectors concatenated

$$\underline{\theta}^{k+1} = \underline{\theta}^{k} - \alpha \nabla_{\underline{\theta}} \ell_{\underline{\theta}}(X_{i(k)}, Y_{i(k)}) \qquad i(k) \sim \mathrm{Uniform}\{1, \dots, N\}$$

→ code