Ex: 1

1. $f(x) = e^{-x}$

$\frac{d}{dx} f(x) = -e^{-x}$ $\quad \frac{d^2}{dx^2} f(x) = e^{-x} > 0 \quad \forall x$ convex

2. $\phi(x) = (\max\{1-x, 0\})^2$

$1-x$ is linear ⇒ convex

$\max\{1-x, 0\}$ is convex because the maximum of a two convex functions are convex.

the range of $\max\{1-x, 0\}$ is ≥ 0 in this range $(\cdot)^2$ is always increasing.

⇒ composition of $(\cdot)^2$ and convex is convex //

3. $\phi(x) = \max\{1-x, 0\}$ is convex as proved before.

4. $1 - \tanh(kx)$ for a fixed x

$$\phi(x) = 1 - \frac{e^{2kx}-1}{e^{2kx}+1}$$

$$\frac{d}{dx}\phi(x) = -\left[\frac{2ke^{2kx}(e^{2kx}+1) - (e^{2kx}-1)2ke^{2kx}}{(e^{2kx}+1)^2}\right]$$

$$= -2ke^{2kx}\left[\frac{2}{(e^{2kx}+1)^2}\right] = \frac{-4ke^{2kx}}{(e^{2kx}+1)^2}$$

$$\frac{d^2}{dx^2}\phi(x) = -\left[\frac{8k^2e^{2kx}}{(e^{2kx}+1)^2} + \frac{-2 \cdot 2ke^{2kx}(4ke^{2kx})}{(e^{2kx}+1)^3}\right]$$

$$= -\frac{8k^2e^{2kx}}{(e^{2kx}+1)^3}\left[e^{2kx}+1-2e^{2kx}\right]$$

when $x = 0$ $\quad < 0$ ⇒ not convex

Ex 2

$$R_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} l(y_i, x_i, \beta) = \frac{1}{n}\sum_{i=1}^{n} \phi(y_i x_i^T \beta)$$

$$= \frac{1}{n}\sum_{i=1}^{n} (1 - y_i x_i^T \beta)_+^2$$

$$\frac{\partial R_n(\beta)}{\partial \beta} = \begin{cases} \frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial \beta}(1 - y_i x_i^T \beta)^2 & 1 - y_i x_i^T \beta > 0 \\ 0 & 1 - y_i x_i^T \beta < 0 \\ \left[0, \frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial \beta}(1 - y_i x_i^T \beta)^2\right] & 1 - y_i x_i^T \beta = 0 \end{cases}$$

sub gradient

$\Rightarrow \frac{\partial R_n(\beta)}{\partial \beta}$ doesn't always exist when $1 - y_i x_i^T \beta = 0$.

with sub gradient,

$$\nabla R_n(\beta) = \begin{cases} -\frac{2}{n}\sum_{i=1}^{n} (1 - y_i x_i^T \beta) y_i \underline{x_i} & 1 - y_i x_i^T \beta > 0 \\ 0 & \text{otherwise} \end{cases}$$

3. with Ridge,

$$R_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} l(y_i x_i^T \beta) + \frac{\lambda}{2}\|\beta\|^2$$

$$\nabla R_n(\beta) = \nabla R_n(\beta) + \lambda\beta$$

$\Rightarrow$

$$\nabla R_n(\beta) = \lambda\beta - \frac{2}{n}\sum_{i=1}^{n} (1 - y_i x_i^T \beta) y_i \underline{x_i} \cdot 1\{1 - y_i x_i^T \beta > 0\}$$

$$= \lambda\beta - \frac{2}{n}\sum_{i=1}^{n} (1 - y_i x_i^T \beta) y_i x_i \, 1\{y_i x_i^T \beta < 1\}$$

(2)

case 1

$$\lim_{(1-y_i x_i^T\beta)\to 0^+} \frac{\partial^2 R(\beta)}{\partial^2 \beta} = +\frac{2}{n}\sum_{i=1}^{n} y_i x_i^T x^{2}$$

$$\lim_{(1-y_i x_i^T\beta)\to 0^-} \frac{\partial^2 R(\beta)}{\partial^2 \beta} = 0$$

$\Rightarrow$ does not continuous.

$$\Rightarrow \nabla R_n(\beta) = \lambda\beta - \frac{2}{n}\sum_{i=1}^{n}(1 - y_i x_i^T\beta)(y_i x_i) \cdot 1\{y_i x_i^T\beta < 1\}$$

4. $\beta \Leftarrow \beta - \alpha^{-1}\nabla L(\beta)$

consider $v \in \mathbb{R}^{p+1}$ $\quad v^T H(\beta) v$

$$= \frac{2}{n}\sum_{i=1}^{n} v^T x_i \, 1\{y_i x_i^T\beta < 1\} x_i^T v + 2\lambda v^T v$$

$$= \frac{2}{n}\sum_{i=1}^{n} \underbrace{(x_i^T v)^T (x_i^T v)}_{\geq 0} \underbrace{1\{y_i x_i^T\beta < 1\}}_{1 \text{ or } 0} + \underbrace{2\lambda v^T v}_{\geq 0 \text{ as } \lambda > 0}$$

$$\Rightarrow \quad v^T H(\beta) v \geq 0 \quad \forall \ v \in \mathbb{R}^{p+1}$$

~~for~~ Pseudo-code

let $\alpha = H(\beta)$, $L(\beta) = R_n(\beta)$

for each iteration:

compute the gradient $\nabla L(\beta)$

$\beta \Leftarrow \beta - \alpha^{-1}\nabla L(\beta)$

stop when loss is small enough.

Complexity: to calculate $H(\beta) = O(\cancel{p^2 n})$ $O(pn)$

to get the inverse of $H(\beta) = p^3$.

Then multiplication $\alpha^{-1}\nabla L(\beta) = O(p^2)$

total $= O(p^3 + p^2 + pn)$ //