1.1) Q4) $f(w) = \frac{1}{2}\underbrace{\sum_{i=1}^{n} z_i (w^T x_i - y_i)^2}_{} + \frac{\lambda}{2}\|w\|^2$

can rewrite this as $\frac{1}{2}(Xw - y)^T Z (Xw - y)$

$= \|Xw - y\|_Z^2$ (conven since it's a norm of a linear function)

$z_i \rightarrow$ weights for the examples like in weighted least squares

or $z_i = \frac{1}{\sigma_i^2}$ [variance for each example is different]

After re-writing in matrix notation, use matrix cook-book

5) use derivative of $c_i = p_i$ with chain rule

c_i ← CDF, p_i ← PDF

→ can use Hadamard product / element-wise product to write the final answer

$a \circ b = \begin{bmatrix} a_1 b_1 \\ a_i b_i \\ \vdots \\ a_n b_n \end{bmatrix}$ $n \times 1$

Hadamard product

1.3) Bayes rule: $\underbrace{p(h|D)}_{\text{MAP}} \propto \underbrace{p(D|h)}_{\text{likelihood}} \underbrace{p(h)}_{\text{prior}}$

$-\log p(h|D) = -\log p(D|h) - \log p(h) + \text{constant}$

if $\underbrace{p(D|h)}_{p(y_i)} = N(w^T x_i, 1)$

$\Rightarrow -\log p(D|h) = \frac{1}{2}\|x_i w - y_i\|_2$

~~To~~ use same arguments for the other parts

2.2) 3) $\|w\|_p \rightarrow$ for showing this is convex, use zeroth-order definition along with triangle inequality

$\hookrightarrow \|x+y\|_p \leq \|x\|_p + \|y\|_p$

other useful things

1) sum of convex functions is convex
2) max " " " is convex
3) $f(Ax)$ is convex if f is convex

$$\nabla f(Ax) = A^T \nabla f(Ax)$$

$$\nabla^2 f(Ax) = A^T \nabla^2 f(Ax) A$$

4) To show logistic regression is convex, show that $\nabla^2 f(x)$ is +ve semi-definite

$$\nabla^2 f(x) = X^T D \cancel{X}$$ [from lecture slides]

$$D_{ii} = \sigma(y_i w^T x_i)\,\sigma(-y_i w^T x_i) \geq 0$$

To show matrix A is +ve semi-definite, show

$$\forall z \quad z^T A z \geq 0$$

2.3

If $f(w) = \sum_{i=1}^{n} \max(a_i, b_i, c_i)$

introduce slack variable $s_i \geq (a_i)$; $s_i \geq b_i$; $s_i \geq c_i$

$$\left.\begin{aligned} \Rightarrow f(w) &= \sum_{i=1}^{n} s_i \\ \text{s.t: } s_i \geq a ;\ & s_i \geq b ;\ s_i \geq c \end{aligned}\right\} \text{ linear Program}$$

use that $|\alpha| = \max(\alpha, -\alpha)$

3.2 Hessian free Newton

$$d = [\nabla^2 f(x)]^{-1} \nabla f(x)$$

$$\Rightarrow [\nabla^2 f(x)] d = \nabla f(x)$$

+ve semi-definite since $f(x)$ is convex

linear system → can be solved using conjugate gradient

(Pcg function in MATLAB)

→ [Need to specify $[\nabla^2 f(x)\, v]$ function]

any vector

$\nabla^2 f(x) = X^T D X$ [for logistic regression]

$[X^T D (X v)]$ → $O(nd)$ operations with $O(nd)$ storage

⇒ Newton method without explicitly storing the Hessian