PE 2014 *

4a).

$$\min_{w,b} \quad \frac{1}{2}\|w\|^2$$

$$g_i(w) = y^{(i)} - w^T x^{(i)} - b - \varepsilon \leq 0$$

$$g_i^*(w) = w^T x^{(i)} + b - y^{(i)} - \varepsilon \leq 0$$

$$\Rightarrow L(w,b,\alpha,\alpha^*) = \frac{1}{2}\|w\|^2 + \sum_{i=1}^m \alpha_i \left(y^{(i)} - w^T x^{(i)} - b - \varepsilon\right)$$

$$+ \sum_{i=1}^m \alpha_i^* \left(w^T x^{(i)} + b - y^{(i)} - \varepsilon\right)$$

$$\cancel{= \frac{1}{2}\|w\|^2 + \sum_{i=1}^m \alpha_i \left(y^{(i)} - w^T x^{(i)}\right)}$$

$$\cancel{+ \sum_{i=1}^m \alpha_i^* \left(w^T x^{(i)} - y^{(i)}\right)}$$

$$\cancel{+ \sum_{i=1}^m \alpha_i \alpha_i^* (}$$

Side note: PE 2014 9 2/2

$$\sum_{i=1}^{m} w^T x^{(i)} (\alpha_i^* - \alpha_i) = \sum_{i=1}^{m} (\alpha_i^* - \alpha_i) \sum_{j=1}^{m} x^{(j)T} (\alpha_j - \alpha_j^*) \cdot x^{(i)}$$

$$= \sum_{i=1}^{m} (\alpha_i^* - \alpha_i) x^{(i)T} \sum_{j=1}^{m} (\alpha_j - \alpha_j^*) x^{(j)}$$

$\Rightarrow \cancel{L(\alpha, \alpha^*)}$ $$= -\sum_{i=1}^{m} (\alpha_i - \alpha_i^*) x^{(i)T} \sum_{i=1}^{m} (\alpha_i - \alpha_i^*) x^{(i)}$$

Side note 2:

$$w^T w = \sum_{i=1}^{m} x^{(i)T} (\alpha_i - \alpha_i^*) \sum_{i=1}^{m} (\alpha_i - \alpha_i^*) x^{(i)}$$

$$\Rightarrow L(\alpha, \alpha^*) = \sum_{i=1}^{m} y^{(i)} (\alpha_i - \alpha_i^*) - 2\varepsilon \sum_{i=1}^{m} \alpha_i$$

$$- \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) x^{(i)T} x^{(j)}$$

$\Rightarrow$ Dual optimization problem!

$$\max_{\alpha, \alpha^*} W(\alpha, \alpha^*) = L(\alpha, \alpha^*) =$$

$$\text{s.t.} \sum_{i=1}^{m} \alpha_i = \sum_{i=1}^{m} \alpha_i^*$$

$$\text{st } \alpha_i, \alpha_i^* \geq 0, \forall i$$

c) Can be kernelized, because inner product.

$\Rightarrow$ Enough to train

to predict:

$$w^T x = \sum_{i=1}^{m} (\alpha_i - \alpha_i^*) x^{(i)T} x$$

inner product

b:

PE 2014 9 1/2 ε

b) $\nabla_w L(w,b,\alpha,\alpha^*) = w - \sum_{i=1}^m \alpha_i \cdot x^{(i)} + \sum_{i=1}^m \alpha_i^* x^{(i)}$

$\nabla_b L(w,b,\alpha,\alpha^*) = \sum_{i=1}^m (\alpha_i^* - \alpha_i) = 0 \Rightarrow \sum_{i=1}^m \alpha_i = \sum_i \alpha_i^*$

$$\Rightarrow \quad w = \sum_{i=1}^m \left[x^{(i)} (\alpha_i - \alpha_i^*) \right] \qquad \mathrm{I}$$

$$\Rightarrow \quad \& \sum_{i=1}^m \alpha_i = \sum_{i=1}^m \alpha_i^* \qquad \mathrm{II}$$

Back in the primal:

$$L(w,b,\alpha,\alpha^*) = \frac{1}{2}\|w\|^2 + \sum_{i=1}^m \Big[\alpha_i^* w^T x^{(i)} - \alpha_i w^T x^{(i)} + y^{(i)} (\alpha_i - \alpha_i^*) \Big]$$

$$\left. + b \sum_{i=1}^m \alpha_i^* - b \sum_{i=1}^m \alpha_i \right\} 0, \text{ because of II}$$

$$\left. - \varepsilon \sum_{i=1}^m \alpha_i - \varepsilon \sum_{i=1}^m \alpha_i^* \right\} 2 \cdot \varepsilon \cdot \sum_{i=1}^m \alpha_i, \ \mathrm{I}$$

$$= \frac{1}{2}\|w\|^2 + \sum_{i=1}^m y^{(i)} (\alpha_i - \alpha_i^*) - 2 \cdot \varepsilon \cdot \sum_{i=1}^m \alpha_i + \sum_{i=1}^m w^T x^{(i)} (\alpha_i^* - \alpha_i)$$

b)

$$\nabla_\theta \ell(\theta) = \sum_{i=1}^{m} x^{(i)} \cdot (y^{(i)} - 1) + \frac{1}{1-\exp(\theta^T x^{(i)})} \cdot (-\exp(\theta^T x^{(i)}) x^{(i)}$$

$$f(g(x)) = f'(g(x)) \cdot g'(x)$$

$$\frac{x^5 \cdot x^{-5}}{x^{-5}}$$

$$\cancel{f} \quad g(\theta) = 1 - \exp(\theta^T x^{(i)})$$

$$g'(\theta) = -\exp(\theta^T x^{(i)}) \cdot x^{(i)}$$

$$\frac{x^5}{1-x^5} = \frac{1}{(1-x^5)(x^{-5})} = \frac{1}{x^{-5}-1} =$$

$$\Rightarrow \nabla_\theta \ell(\theta) = \sum_{i=1}^{m} x^{(i)} \cdot (y^{(i)} - 1) - \frac{x^{(i)} \cdot \exp(\theta^T x^{(i)})}{1 - \exp(\theta^T x^{(i)})}$$

$$= \sum_{i=1}^{m} x^{(i)} \cdot (y^{(i)} - 1) - \frac{x^{(i)}}{\exp(-\theta^T x^{(i)}) - 1}$$

$$\cancel{x^{(i)} (y^{(i)} - 1 - 1)}$$

For Hesse:

$$\frac{\nabla \ell(\theta)}{d\theta_i \theta_j} = \sum$$

$$((\exp(-\theta^T x^{(i)} - 1)^{-1})'$$

$$= -\exp(-\theta^T x^{(i)} - 1)^{-2} \cdot -x_j^{(i)}$$

$$= x_j^{(i)} \cdot (\exp(-\theta^T x^{(i)}) - 1)^{-2}$$

$$\cancel{Ab} \; (\cancel{\tfrac{1}{x}} (x^{-1})' = -1 \cdot \frac{1}{x^2}$$