CS6140 Assignment 4

2.3 $$LL_{Regularized} = \frac{\lambda}{2}||w||^2 + \sum_{i=1}^{N}\left\{\ln\left(1+e^{-y_i w^T x_i}\right)\right\}$$

$$= \frac{\lambda}{2}||w||^2 + \sum_{i=1}^{N}\left\{y_i \cdot \ln O_i + (1-y_i)\ln(1-O_i)\right\}$$

$$\frac{\partial LL}{\partial w} = \lambda||w|| + \sum_{i=1}^{N}\left\{y_i(1-O_i) \times x_i \cdot (1-y_i)(-1) \times O_i x_i\right\}$$

$$= \lambda||w|| + \sum_{i=1}^{N}\left\{y_i x_i - \cancel{y_i O_i x_i} - O_i x_i + \cancel{y_i O_i x_i}\right\}$$

$$= \lambda||w|| + \sum_{i=1}^{N}(y_i - O_i)x_i$$

learning rate

$$w_0 = w_0 - l\sum_{i=1}^{N}(y_i - O_i)x$$

$$w_r = w_r + \frac{\lambda}{c}w_r - l\sum_{i=1}^{N}(y_i - O_i)x$$

Gradient Descent Algorithm

(2.5) Yes, it is possible to kernelize equation 2.

$$LL(\omega) = \sum_{i=1}^{N} \ln(\text{sigmoid}(y_i \omega^T x_i)) - \frac{\lambda}{2} \|\omega\|^2$$

$$\ln(\text{sigmoid}(x)) \leq \alpha x - H(\alpha)$$

$$\text{where } H(\alpha) = -\alpha \log \alpha - (1-\alpha)\log(1-\alpha)$$

$$\rightarrow LL(\omega) \leq LL(\omega, \alpha)$$

$$\text{where } LL(\omega,\alpha) = \sum_{i=1}^{N} \alpha_i y_i \omega^T x_i - H(\alpha_i) - \frac{\lambda}{2} \|\omega\|^2$$

$$\& \ \omega(\alpha) = \lambda^{-1} \sum_{i=1}^{n} \alpha_i y_i x_i$$

$$\boxed{LL(\alpha) = \frac{1}{2\lambda} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j x_j^T x_i - \sum_{i=1}^{N} H(\alpha_i)}$$

Dual Form

Solution learned from "A comparison of numerical optimizers for logistic regression" by Thomas P. Minka 10/22/2003
and provided Andrew Ng video.