EE 5841 Machine Learning

Project 4

Kernel Logistic Regression

Derivation of the dual formulation for linear logistic regression:

Let (x_i, y_i) denote the training set where $x_i \in R^d$ and $y_i \in$ (-1,1)

$$P(y|x) = \frac{1}{1 + \exp\left(-y_i(w^T.\mathrm{x} + \mathrm{b})\right)} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (1)$$

Taking max log likelihood and adding the regularization term

$$\min E = \frac{\lambda}{2} \|\mathrm{w}\|^2 + \sum_i \log(1 + e^{\xi}) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (2)$$

Where $\xi = -y_i(w^T\mathrm{x} + \mathrm{b}) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (3)$

Taking $g(\xi) = \log\left(1 + e^{\xi}\right) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (4)$

Taking Lagrangian

$$\max L = \frac{\lambda}{2} \|\mathrm{w}\|^2 + \sum_i g(\xi_i) + \sum_i \alpha_i(-\xi_i - y_i(w.x_i - b)) \dots\dots\dots\dots\dots (5)$$

To obtain the equation in terms of α

Taking Partial derivative of the Lagrangian equation w.r.t. w

$$\nabla_w L = \lambda w - \sum_i \alpha_i y_i x_i = 0 \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (6)$$

Taking Partial derivative of the Lagrangian equation w.r.t. b

$$\frac{\partial L}{\partial b} = \sum_i \alpha_i y_i = 0 \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (7)$$

Taking Partial derivative of the Lagrangian equation w.r.t. ξ_i

$$\frac{\partial L}{\partial \xi} = g'(\xi_i) - \alpha_i = 0$$

$$\xi_i = {g'}^{-1}(\alpha_i) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (8)$$

From equation (5)

$$\max L = \frac{\lambda}{2}\|\mathrm{w}\|^2 + \sum_i g(\xi_i) - \sum_i \alpha_i \xi_i - \sum_i (\alpha_i y_i x_i).w + \sum_i \alpha_i y_i b \ldots\ldots\ldots\ldots (9)$$

Taking $G(\alpha_i) = \alpha_i \xi_i - g(\xi_i) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. (10)$

Taking partial derivative of Equation (9) w.r.t. α_i

$$\frac{\partial G(\alpha_i)}{\partial \alpha} = (\xi_i \alpha_i \frac{\partial \xi}{\partial \alpha} + (1)\xi_i) - g'(\xi_i)\frac{\partial \xi}{\partial \alpha}$$

$$= \xi_i + \alpha_i \frac{\partial \xi}{\partial \alpha} - g'(\xi_i)\frac{\partial \xi}{\partial \alpha}$$

Substituting equation (8)

$$= \xi_i + \alpha_i \frac{\partial \xi}{\partial \alpha} - \alpha_i \frac{\partial \xi}{\partial \alpha}$$

$$= \xi_i$$

Substituting equation (8)

$$= g'^{-1}(\alpha_i)$$

$g'(\alpha_i)$ is a convex function, therefore $g'^{-1}(\alpha_i)$ is also a convex function

$$\frac{\partial G(\alpha_i)}{\partial \alpha} = G'(\alpha_i) = \log\frac{\alpha_i}{1-\alpha_i} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots. (11)$$

Therefore,

$$\xi_i = \log\frac{\alpha_i}{1-\alpha_i} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots. (12)$$

From equation (4)

$$g(\xi) = \log\left(1 + e^{\xi}\right)$$

Substituting equation (12)

$$g(\xi) = \log\left(1 + e^{\log_e \frac{\alpha_i}{1-\alpha_i}}\right)$$

$$g(\xi) = \log\left(1 + \frac{\alpha_i}{1 - \alpha_i}\right) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots. (13)$$

From equation (9)

$$\max L = \frac{\lambda}{2}\|\mathrm{w}\|^2 + \sum_i g(\xi_i) - \sum_i \alpha_i \xi_i - \sum_i (\alpha_i y_i x_i).w + \sum_i \alpha_i y_i b$$

Substituting equation (10)

$$\max L = \frac{\lambda}{2}\|\mathrm{w}\|^2 - \sum_i G(\alpha_i) - \sum_i (\alpha_i y_i x_i).w + \sum_i \alpha_i y_i b \dots\dots\dots\dots\dots\dots\dots\dots\dots (14)$$

Substituting equation (7)

$$\max L = \frac{\lambda}{2}\|\mathrm{w}\|^2 - \sum_i G(\alpha_i) - \sum_i (\alpha_i y_i x_i).w$$

Substituting equation (6)

$$\max L = \frac{\lambda}{2}\|\mathrm{w}\|^2 - \sum_i G(\alpha_i) - \lambda w^T.w$$

$$\max L = \frac{\lambda}{2}\|\mathrm{w}\|^2 - \sum_i G(\alpha_i) - \lambda\|\mathrm{w}\|^2$$

$$\max L = -\frac{\lambda}{2}\|\mathrm{w}\|^2 - \sum_i G(\alpha_i)$$

To maximize the above equation, we have to minimize the function of α

Therefore, the Dual formulation of the Linear Logistic Regression will be

$$\min f(\alpha) = -\frac{\lambda}{2}\|\mathrm{w}\|^2 + \sum_i G(\alpha_i) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (15)$$

Subject to $\sum_i \alpha_i y_i = 0\,, \forall i$

Kernel Version of the Dual formation of logistic regression:

From equation (15)

$$\min f(\alpha) = \frac{\lambda}{2}\|\mathrm{w}\|^2 + \sum_i G(\alpha_i)$$

Substituting equation (6)

$$\min f(\alpha) = \frac{\lambda}{2}(|\frac{1}{\lambda}\sum_i \alpha_i y_i x_i\ |^T . |\frac{1}{\lambda}\sum_j \alpha_j y_j x_j\ |) + \sum_i G(\alpha_i)$$

$$\min f(\alpha) = \frac{1}{2\lambda}\sum_i \sum_j \alpha_i \alpha_j y_i y_j (x_i).(x_j) + \sum_i G(\alpha_i) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots . (16)$$

Subject to $\sum_i \alpha_i y_i = 0 , \forall i$

Where $(x_i).(x_j)$ is the Kernel Term

Kernel Version of Pr(y|x) used the obtained dual variables:

From equation (1)

$$P(y|x) = \frac{1}{1 + \exp\left(-y_i\left(w^T . x_j + \mathrm{b}\right)\right)}$$

Substituting equation (6)

$$P(y|x) = \frac{1}{1 + \exp\left(-\mathrm{y}\left(\frac{1}{\lambda}\sum_i \alpha_i y_i (x_i) . (x_j) + \mathrm{b}\right)\right)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (17)$$

Subject to $\sum_i \alpha_i y_i = 0 , \forall i$

Where $(x_i).(x_j)$ is the Kernel Term