MACHINE LEARNING 1

Bui Hoang Linh - 11205706

Homework 5

1 Problem 1

Calculate vector calculus $\frac{\partial L}{\partial w} = X^T(\hat{y} - y)$

$$L = -log\ p(t|w) = -\sum_{i=1}^{N} y_i log(\hat{y_i}) + (1 - y_i) log(1 - \hat{y_i})$$
$$= -(y\ log\hat{y} + (1 - y) log(1 - \hat{y}))$$

$$\hat{y} = \sigma(X^T w) = \frac{1}{1 + e^{-X^T w}}\ ,\ z = e^{-X^T w}$$

We have:

$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial \hat{y}} . \frac{\partial \hat{y}}{\partial z} . \frac{\partial z}{\partial w}$ (chain rule)

- $\frac{\partial L}{\partial \hat{y}} = -\Big(y . \frac{1}{\hat{y}} - (1 - y) . \frac{1}{1 - \hat{y}}\Big) = -\Big(\frac{y}{\hat{y}} - \frac{1 - y}{1 - \hat{y}}\Big)$
- $\frac{\partial \hat{y}}{\partial w} = \frac{\partial \hat{y}}{\partial z} . \frac{\partial z}{\partial w}$
  $= -\frac{1}{(1 + z)^2} . (-Xe^{-X^T w}) = X . \frac{1}{1 + z} . \frac{z}{1 + z} = X\hat{y}(1 - \hat{y})$

$$\Rightarrow \frac{\partial L}{\partial w} = -\Big(\frac{y}{\hat{y}} - \frac{1 - y}{1 - \hat{y}}\Big) . X\hat{y}(1 - \hat{y})$$
$$= \frac{-y + y\hat{y} + \hat{y} - y\hat{y}}{\hat{y}(1 - \hat{y})} . X\hat{y}(1 - \hat{y})$$
$$= X(\hat{y} - y)$$

Under the matrix form: $\frac{\partial L}{\partial w} = X^T(\hat{y} - y)$