Linear regression

$y_i = w^T \bar{x}_i$

$$E(w) = \frac{1}{2}\sum_i \left(w^T\bar{x}_i - t_i\right)^2 = \frac{1}{2}\sum_i \left(y_i - t_i\right)^2 =$$

$$= \frac{1}{2}(y-t)^T(y-t)$$

$$\underbrace{\begin{bmatrix} y_1-t_1 & y_2-t_2 & \cdots & y_n-t_n \end{bmatrix}}_{(y-t)^T} \underbrace{\begin{bmatrix} y_1-t_1 \\ \vdots \\ y_n-t_n \end{bmatrix}}_{y-t} = \sum_i (y_i-t_i)^2$$

$$n\;\underbrace{\begin{array}{c} d+1 \\ \begin{bmatrix} - \; \bar{x}_1 \; - \\ - \; \bar{x}_2 \; - \\ \vdots \\ - \; \bar{x}_n \; - \end{bmatrix} \end{array}}_{\bar{X}} \underbrace{\begin{array}{c} 1 \\ \begin{bmatrix} w_0 \\ \vdots \\ w_d \end{bmatrix} \end{array}}_{w} \, d+1 = \underbrace{\begin{array}{c} 1 \\ \begin{bmatrix} w^T\bar{x}_1 \\ \vdots \\ w^T\bar{x}_n \end{bmatrix} \end{array}}_{\bar{X}w} \, n = \underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}}_{y}$$

$$E(w) = \frac{1}{2}\left(\bar{X}w - t\right)^T\left(\bar{X}w - t\right)$$

$$\frac{\partial E(w)}{\partial w} = \sum_i (y_i - t_i)\bar{x}_i = (y_1-t_1)\begin{bmatrix} \bar{x}_1 \end{bmatrix} + \dots + (y_n - t_n)\begin{bmatrix} \bar{x}_n \end{bmatrix} =$$

$$= d+1\;\underbrace{\begin{array}{c} n \\ \begin{bmatrix} | & | & & | \\ \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_n \\ | & | & & | \end{bmatrix} \end{array}}_{\bar{X}^T} \underbrace{\begin{array}{c} 1 \\ \begin{bmatrix} y_1 - t_1 \\ y_2 - t_2 \\ \vdots \\ y_n - t_n \end{bmatrix} \end{array}}_{(y-t)} n \;=\; \bar{X}^T(y-t) = \bar{X}^T\left(\bar{X}w - t\right)$$

$$H = \sum_i \bar{x}_i x_i^T = \sum_i \begin{array}{c} d+1 \\ \begin{bmatrix} x_{i0}^2 & x_{i0}x_{i1} & \cdots & x_{i0}x_{id} \\ x_{i0}x_{i1} & x_{i1}^2 & \cdots & x_{i1}x_{id} \\ \vdots & \vdots & \ddots & \vdots \\ x_{i0}x_{id} & x_{i1}x_{id} & \cdots & x_{id}^2 \end{bmatrix} \end{array} d+1 =$$

$$= \begin{bmatrix} \sum_i x_{i0}^2 & \sum_i x_{i0}x_{i1} & \cdots & \sum x_{i0}x_{id} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_i x_{i0}x_{id} & \sum_i x_{i1}x_{id} & \cdots & \sum x_{id}^2 \end{bmatrix} =$$

$$= \begin{array}{c c} & n \\ d+1 & \underbrace{\begin{bmatrix} | & | & & | \\ \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_n \\ | & | & & | \end{bmatrix}}_{\bar{X}^T} \end{array} \begin{array}{c c} d+1 & \\ \underbrace{\begin{bmatrix} - & \bar{x}_1 & - \\ - & \bar{x}_2 & - \\ & \vdots & \\ - & \bar{x}_n & - \end{bmatrix}}_{\bar{X}} & n \end{array} = \bar{X}^T \bar{X}$$

$$w^{(i+1)} = w^{(i)} - (\bar{X}^T \bar{X})^{-1} \bar{X}^T (y - t) =$$

$$y = \bar{X} w$$

$$= w^{(i)} - (\bar{X}^T \bar{X})^{-1} \bar{X}^T y + (\bar{X}^T \bar{X})^{-1} \bar{X}^T t =$$

$$= w^{(i)} - (\bar{X}^T \bar{X})^{-1} \bar{X}^T \bar{X} w^{(i)} + (\bar{X}^T \bar{X})^{-1} \bar{X}^T t =$$

$$= w^{(i)} - w^{(i)} + (\bar{X}^T \bar{X})^{-1} \bar{X}^T t =$$

$$= (\bar{X}^T \bar{X})^{-1} \bar{X}^T t$$

Logistic regression

$a_i = w^T \bar{x}_i$

$y_i = \sigma(a_i)$

$$E(w) = -\sum_i \left(t_i \lg y_i + (1-t_i)\lg(1-y_i) \right)$$

$$\frac{\partial E(w)}{\partial w} = \sum_i (y_i - t_i)\bar{x}_i = \bar{X}^T (y - t) =$$

$$H = \sum_i y_i(1-y_i)\bar{x}_i \bar{x}_i^T = \sum_i \begin{bmatrix} x_{i0}^2 & x_{i0}x_{i1} & \cdots & x_{i0}x_{id} \\ \vdots & \vdots & & \vdots \\ x_{i0}x_{id} & x_{i1}x_{id} & & x_{id}^2 \end{bmatrix} y_i(1-y_i)$$

$$= \begin{bmatrix} \sum_i x_{i0}^2 y_i(1-y_i) & \sum_i x_{i0}x_{i1} y_i(1-y_i) & \cdots & \sum_i x_{i0}x_{id} y_i(1-y_i) \\ \\ \sum_i x_{i0}x_{id} y_i(1-y_i) & \sum_i x_{i1}x_{id} y_i(1-y_i) & \cdots & \sum_i x_{id}^2 y_i(1-y_i) \end{bmatrix} =$$

$$= \underbrace{\begin{array}{c|c} & n \\ \hline d+1 & \begin{bmatrix} | & | & & | \\ \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_n \\ | & | & & | \end{bmatrix} \end{array}}_{\bar{X}} \underbrace{\begin{array}{c|c} & n \\ \hline n & \begin{bmatrix} y_1(1-y_1) & & 0 \\ & y_2(1-y_2) & \\ 0 & & \ddots \; y_n(1-y_n) \end{bmatrix} \end{array}}_{Y} \underbrace{\begin{array}{c|c} d+1 & \\ \hline \begin{bmatrix} - & \bar{x}_1 & - \\ - & \bar{x}_2 & - \\ & \vdots & \\ - & \bar{x}_n & - \end{bmatrix} & n \end{array}}_{X}$$

$$w^{(i+1)} = w^{(i)} - (\bar{X}^T Y \bar{X})^{-1} \bar{X}^T (y-t) =$$

$$= \boxed{(\bar{X}^T Y \bar{X})^{-1} \bar{X}^T Y \bar{X}} \; w^{(i)} - \boxed{(\bar{X}^T Y \bar{X})^{-1} \bar{X}^T Y} Y^{-1} (y-t)$$

$$= (\bar{X}^T Y \bar{X})^{-1} \bar{X}^T Y \left(\bar{X} w^{(i)} - Y^{-1} (y-t) \right) =$$

$$= (\bar{X}^T Y \bar{X})^{-1} \bar{X}^T Y \left(a - Y^{-1} (y-t) \right) =$$

$$= (\bar{X}^T Y \bar{X})^{-1} \bar{X}^T Y \, z$$

$$\underbrace{\begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}}_{z} = \underbrace{\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}}_{a} - \underbrace{\begin{bmatrix} \frac{1}{y_1(1-y_1)} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{y_n(1-y_n)} \end{bmatrix}}_{Y^{-1}} \underbrace{\begin{bmatrix} y_1 - t_1 \\ \vdots \\ y_n - t_n \end{bmatrix}}_{y-t}$$