Least square fit

- $E(\boldsymbol{w}) = (\boldsymbol{X}\boldsymbol{w} - \boldsymbol{Y})^T(\boldsymbol{X}\boldsymbol{w} - \boldsymbol{Y})$

$$\text{where } \boldsymbol{X} = \begin{bmatrix} 1 & x_1 & \cdots & {x_1}^M \\ \vdots & & \ddots & \vdots \\ 1 & x_N & \cdots & {x_N}^M \end{bmatrix}, \boldsymbol{w} = \begin{bmatrix} w_0 \\ \vdots \\ w_m \end{bmatrix}, \boldsymbol{Y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}$$

- $\boldsymbol{w} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$

Regularized least square fit

$$\boldsymbol{w} = (\lambda \boldsymbol{I} + \boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$$

Gradient decent for single variable linear regression

- $\boldsymbol{w}^{(i+1)} = \boldsymbol{w}^{(i)} - \alpha \frac{dE(\boldsymbol{w})}{d\boldsymbol{w}}\Big|_{\boldsymbol{w}=\boldsymbol{w}_i} = \boldsymbol{w}^{(i)} - 2\alpha(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{w}^{(i)} - \boldsymbol{X}^T\boldsymbol{Y})$
- $\frac{dE(\boldsymbol{w})}{d\boldsymbol{w}} = 2(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{w} - \boldsymbol{X}^T\boldsymbol{Y}) = 2\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{w} - \boldsymbol{Y})$

$$= 2\begin{bmatrix} \sum_{i=1}^{N}(w_0 + w_1x_i - y_i) \\ \sum_{i=1}^{N}(w_0 + w_1x_i - y_i)x_i \end{bmatrix}$$

Logistic regression

- $\boldsymbol{w}^{(i+1)} = \boldsymbol{w}^{(i)} - \alpha \frac{dC(\boldsymbol{w})}{d\boldsymbol{w}}\Big|_{\boldsymbol{w}=\boldsymbol{w}^{(i)}} = \boldsymbol{w}^{(i)} + \alpha \sum_{n=1}^{N}(y_n - \hat{p}_n) \cdot \mathrm{x}_n$

Linear discriminant analysis

$$\mathbf{S}_W = \sum_{n\in c_1}(\mathbf{x}_n - \mathbf{m}_1)(\mathbf{x}_n - \mathbf{m}_1)^T + \sum_{n\in c_2}(\mathbf{x}_n - \mathbf{m}_2)(\mathbf{x}_n - \mathbf{m}_2)^T$$

$$\mathbf{S}_B = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T$$

$\mathbf{w}$ is the eigenvector of $\mathbf{S}_w^{-1}\mathbf{S}_B$

Artificial Neural Network

- Sigmoid: $\sigma(a) = \frac{1}{1+e^{-a}}$
- Tanh: $\tanh(a) = \frac{e^{a}-e^{-a}}{e^{a}+e^{-a}}$
- ReLU (Rectified Linear Unit): $\mathrm{ReLU}(a) = \max(0, a)$
- Step function: $s(a) = \begin{cases} 1 \text{ if a } \geq 0 \\ 0 \text{ if a} < 0 \end{cases}$

Responsibilities of output-layer neurons:

$$\delta_i^{(1)} = y_i(1 - y_i)(t_i - y_i)$$

Responsibilities of hidden-layer neurons:

$$\delta_j^{(2)} = h_j(1 - h_j) \sum_i \delta_i^{(1)} w_{ji}$$

Weight updates:

$$\begin{aligned} &\text{output layer: } w_{ji}^{(1)} := w_{ji}^{(1)} + \eta\delta_i^{(1)} h_j; \\ &\text{hidden layer: } w_{ki}^{(2)} := w_{ki}^{(2)} + \eta\delta_i^{(2)} x_k; \end{aligned}$$