linear regression

Training data

	$1, 2, \dots, j \dots p$	
1		
2		
$\vdots$		
i	$x_{i1} \cdots x_{ij} \dots x_{ip}$	y_i
$\vdots$		
n		

supervised learning

input $(x_{i1} \cdots x_{ij} \cdots x_{ip})$

output y_i supervision

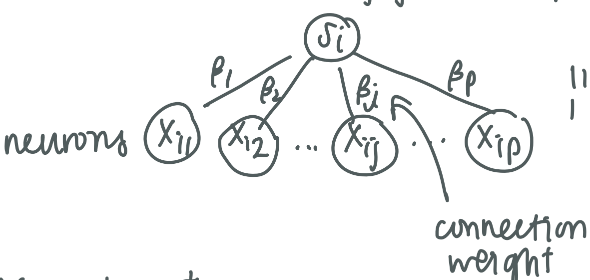
want to learn mapping to apply to other testing situations

input $\longrightarrow$ output

Model:

$$y_i = x_{i1}\beta_1 + \dots + x_{ij}\beta_j + \dots x_{ip}\beta_p + e_i$$

$$s_i = x_{i1}\beta_1 + \dots + x_{ij}\beta_j + \dots + x_{ip}\beta_p$$



linear model: 1 layer NN

intercept/bias

$$s_i = \beta_0 + x_{i1}\beta_1 + \dots x_{ij}\beta_j + \dots + x_{ip}\beta_p$$

$$= \boxed{x_{i0}}\beta_0 + x_{i1}\beta_1 + \dots$$

$$x_{i0} = 1$$

Row vector treatment

$$x_i = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{ij} \\ \vdots \\ x_{ip} \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_j \\ \vdots \\ \beta_p \end{pmatrix}$$

	$1 \dots p$	
1		
i	$— x_i^T —$	y_i row i
$\vdots$		
n		

$$s_i = \langle x_i, \beta \rangle = x_i^T \beta$$

$$\left(= \langle x_i, w \rangle + b \right)$$

w ↓ weight, b ↓ β_0

$$e_i = y_i - s_i$$

$$\text{Loss}(\beta) = \frac{1}{2}\sum_{i=1}^{n} e_i^2$$ least squares loss function

$$\frac{d\text{Loss}}{d\beta_k} = \sum_{i=1}^{n} e_i \frac{de_i}{ds_i}\frac{ds_i}{d\beta_k}$$

simple error back propogation chain rule

$$= -\sum_{i=1}^{n} e_i \cdot x_{ik} \quad \forall k = 1, \dots p$$

$$\text{Loss}'(\beta) = \begin{array}{c} \\ \left(\frac{d\text{Loss}}{d\beta_k} \right) \end{array} \begin{array}{c} 1 \\ \vdots \\ k \\ \vdots \\ p \end{array} = -\sum_{i=1}^{n} \left(x_{ik} e_i \right) \begin{array}{c} 1 \\ \vdots \\ k \\ \vdots \\ p \end{array} = -\sum_{i=1}^{n} \left(x_{ik} \right) \begin{array}{c} 1 \\ \vdots \\ k \\ \vdots \\ p \end{array} e_i = -\sum_{i=1}^{n} \underset{p\times 1}{x_i} \; \underset{1\times 1}{e_i}$$

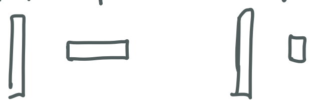
Estimating eqn

$$\sum_{i=1}^{n} x_i(y_i - x_i^T\beta) = 0$$

$$\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \boxed{x_i^T\beta} = 0$$

$$\boxed{\overset{p\times p}{\sum_{i=1}^{n} x_i x_i^T}}\ \beta_{p\times 1} = \boxed{\overset{p\times 1}{\sum_{i=1}^{n} x_i y_i}}$$

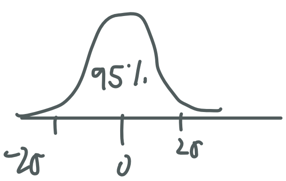
$p\times 1$ $\quad 1\times p$ $\quad p\times 1$ $\quad 1\times 1$

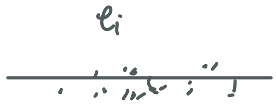


$$\hat{\beta}_{p\times 1} = \underbrace{\left(\sum_{i=1}^{n} x_i x_i^T\right)^{-1}}_{p\times p} \underbrace{\left(\sum_{i=1}^{n} x_i y_i\right)}_{p\times 1}$$

Maximum likelihood

$$e_i \sim N(0, \sigma^2)$$



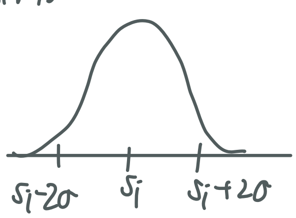


scatter plot

$$[y_i | s_i] \sim N(s_i, \sigma^2)$$

$$y_i = s_i + e_i$$

$$p(y_i|s_i) = \sqrt{2\pi\sigma^2}\ e^{-\frac{(y_i - s_i)^2}{2\sigma^2}}$$ prob density



$x_i^T\beta$

$$\text{likelihood}(\beta) = \prod_{i=1}^{n} p(y_i|s_i)$$

↓ max

most plausible β

$$\text{log-likelihood}(\beta) = \sum_{i=1}^{n} \log p(y_i|s_i)$$

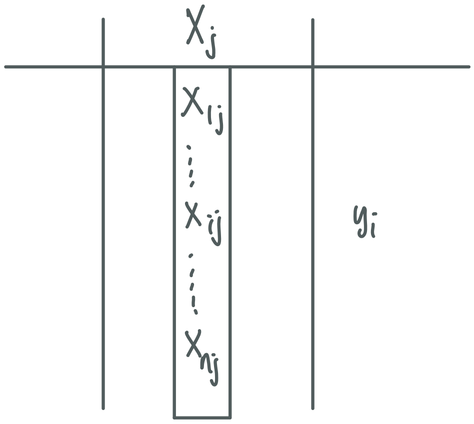
$$= \sum_{i=1}^{n}\left[-\frac{(y_i - s_i)^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right]$$

$$= \frac{-1}{2\sigma^2}\sum_{i=1}^{n}(y_i - s_i)^2 - \frac{n}{2}\log(2\pi\sigma^2)$$

↓ max

$$\min \sum_{i=1}^{n}(y_i - s_i)^2$$ least squares objective function

column vector view



$$X_{n\times p} = (X_1 \cdots X_j \cdots X_p)$$

$$\begin{array}{c} 1 \\ \vdots \\ i \\ \vdots \\ n \end{array} \begin{pmatrix} y_i \end{pmatrix} = \begin{pmatrix} \underbrace{x_{i1}\beta_1 + \dots x_{ij}\beta_j + \dots x_{ip}\beta_p}_{s_i} \end{pmatrix} \begin{array}{c} 1 \\ \vdots \\ i \\ \vdots \\ n \end{array} + \begin{pmatrix} e_i \end{pmatrix} \begin{array}{c} 1 \\ \vdots \\ i \\ \vdots \\ n \end{array}$$

$$Y_{n\times 1} = X_1\beta_1 + \dots + X_j\beta_j + \dots + X_p\beta_p + e$$

pretend X_j as scalars

$$= (X_1 \dots X_j \dots X_p)\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_j \\ \vdots \\ \beta_p \end{pmatrix} + e_{n\times 1}$$

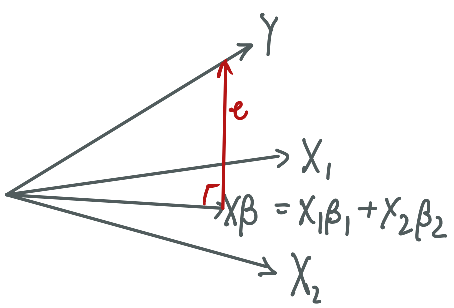
reduce X_j as vector

$$= \underset{n\times p}{X}\ \underset{p\times 1}{\beta} + \underset{n\times 1}{e}$$

$$\text{Loss}(\beta) = \frac{1}{2}\sum_{i=1}^{n} e_i^2 = \frac{1}{2}|e|^2$$

$$\frac{\partial \text{Loss}}{\partial \beta_k} = -\sum_{i=1}^{n} e_i \cdot x_{ik}$$

$$= -\langle e, X_k \rangle \quad k = 1, \dots, p$$



$$X_k^T e = 0 \qquad \forall k = 1, \dots p$$

$$\begin{array}{c} 1 \\ \vdots \\ k \\ \vdots \\ p \end{array} \begin{pmatrix} X_k^T \end{pmatrix} e = 0 \qquad \begin{array}{c} 1 \\ \vdots \\ k \\ \vdots \\ p \end{array} \underbrace{\begin{pmatrix} \boxed{\quad X_k^T \quad} \end{pmatrix}}_{\substack{X^T \\ p\times n}} \begin{bmatrix} e \end{bmatrix}$$

$$X^T(Y - X\beta) = 0$$

$$X^TY - X^TX\beta = 0$$

$$X^TX\beta = X^TY \qquad \hat{\beta} = (X^TX)^{-1}(X^TY)$$

eq. for column vector view

rewrite/revisit row vector view

$$\underbrace{\begin{pmatrix} & x_i & \end{pmatrix}}_{1 \cdots i \cdots n} \begin{pmatrix} x_i^T \end{pmatrix} = \sum_{i=1}^{n} x_i x_i^T$$

$$X^T \quad X$$

$$X^T \quad Y = \sum_{i=1}^{n} x_i y_i$$

$$\hat{\beta} = \left(\sum_{i=1}^{n} x_i x_i^T\right)^{-1} \left(\sum_{i=1}^{n} x_i y_i\right)$$ eq for row vector view

Logistic regression

	$1 \cdots j \cdots p$	
1		
$\vdots$		
i	$x_{i1} \cdots x_{ij} \cdots x_{ip}$	$y_i \in \{0, 1\}$
		$\{-, +\}$
$\vdots$		
n		

$s_i = x_{i1}\beta_1 + \ldots + x_{ij}\beta_j + \ldots + x_{ip}\beta_p$

$= x_i^T \beta$

Extra

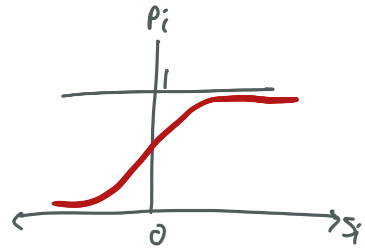
$y_i = \text{sign}(s_i)$ perception

soften

$p_i = P_r(y_i = 1 \mid s_i)$

$= \text{sigmoid}(s_i)$

$= \dfrac{e^{s_i}}{1 + e^{s_i}}$



* least squares

$\text{Loss}(\beta) = \sum_{i=1}^{n} (y_i - \text{sigmoid}(s_i))^2$

✓ maximum likelihood

$\text{log-likelihood}(\beta) = \sum_{i=1}^{n} \log p(y_i \mid s_i)$