Cross validation

K-fold

Grid search   CV

$\left.\begin{array}{c} \end{array}\right\}$ CR

Regularisation

Loss func $\Rightarrow$

$$(y - \hat{y})^2$$

$$\underbrace{(y - (w_1 x_1 + w_2 x_2 + c))^2}_{\underset{0}{ll}} + \lambda w_i^2 \quad \overset{L2}{\underset{\uparrow}{}}$$

$$+ \lambda |w_i|$$

Training set $\Rightarrow$ model

$\lambda \Rightarrow 0.01$

$\lambda = 0.1$

test $\rightarrow R^2$ $x$

test $\rightarrow R^2 \rightarrow y$

$y > x$

data

train (60%)

CV (20%)

test (20%) $\lambda_1$



Error or loss

CV-loss

Train loss

overfit

$\lambda_{best}$

[00%

80%
Train

20%
CU

CU

Train

20%  20%  20%  20%  20%

| Test | Train | Trai | Train | Train | $\rightarrow R_1^2$ |

| Train | Test | T | T | T | $\rightarrow R_2^2$ |

| | | | | | $\rightarrow R_3^2$ |

| | | | | | $\rightarrow R_4^2$ |

$\rightarrow R_S^2$
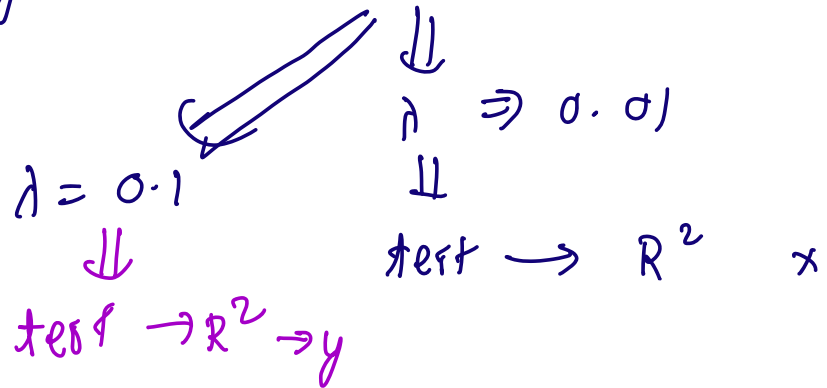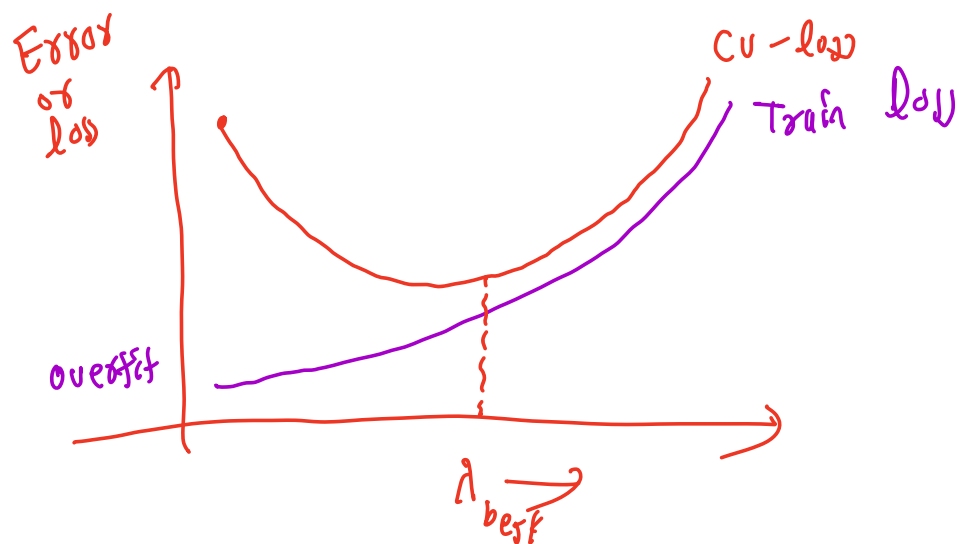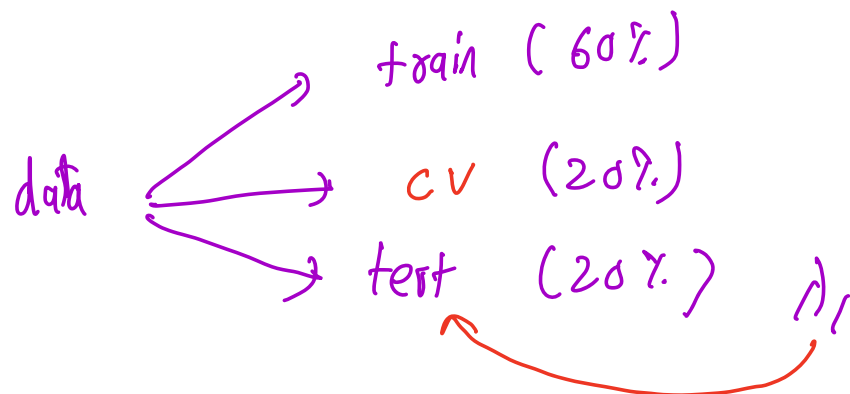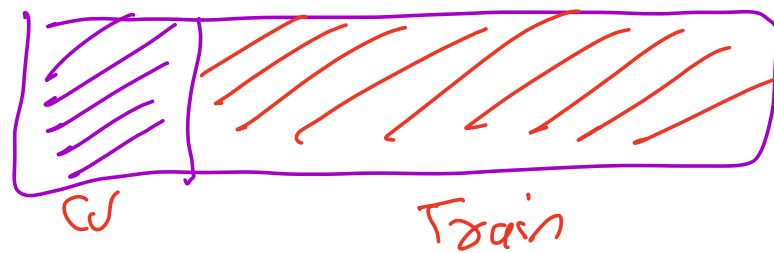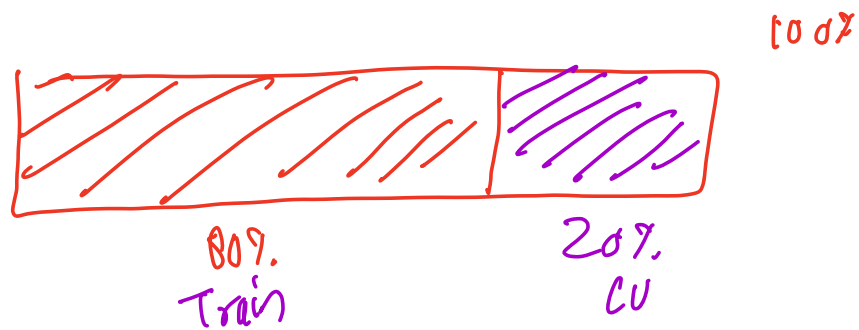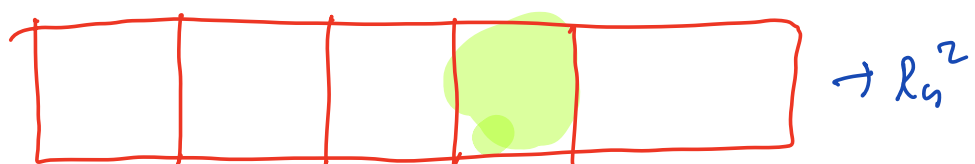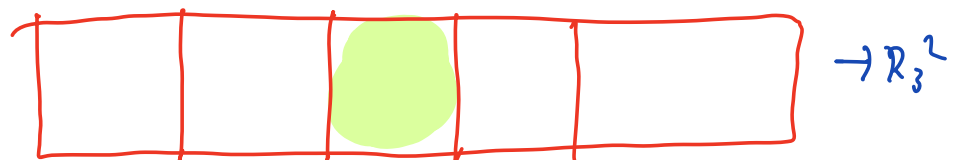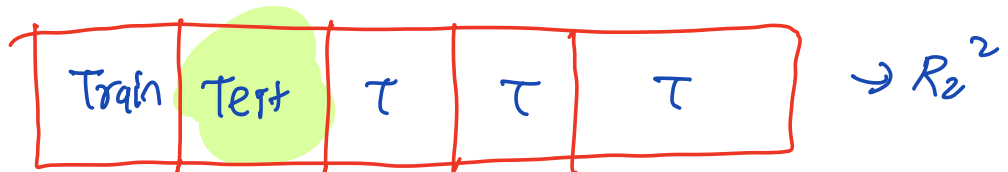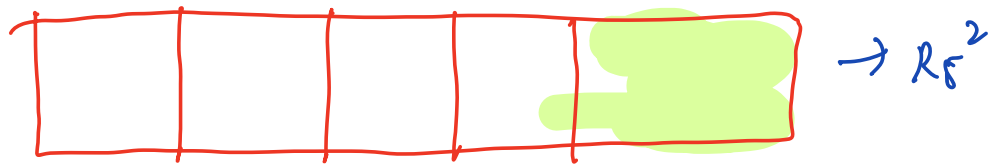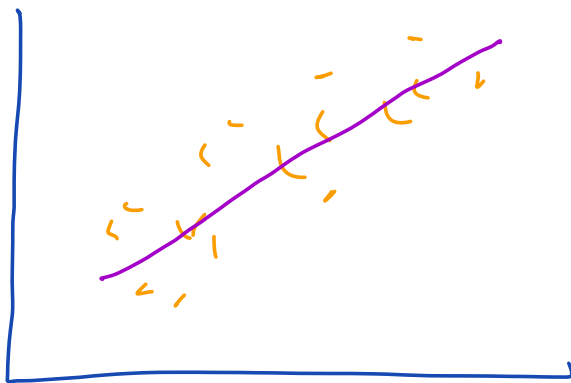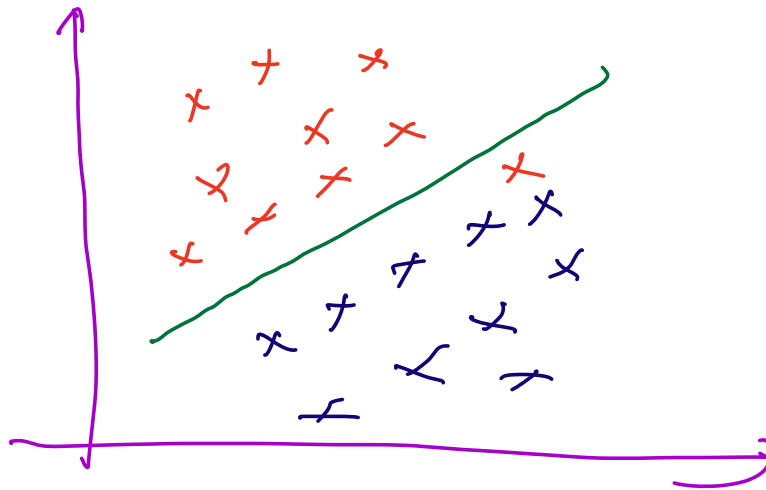
Break: 10: 30 PM

## Logistic Regression

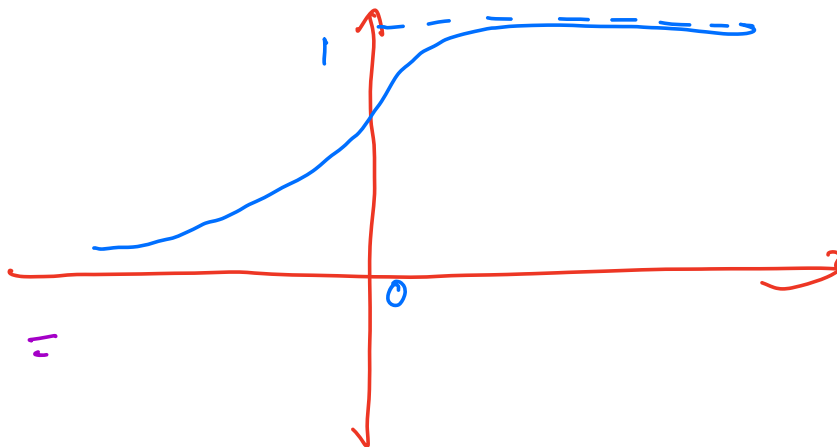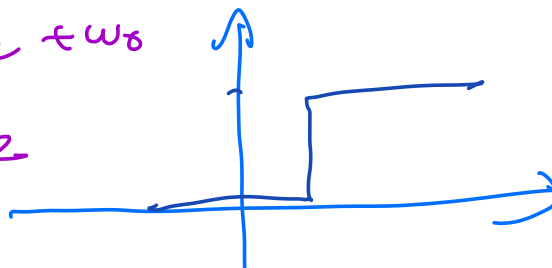Real number
bingay classification

$$y_i \in \{0, 1\}$$

$$w_1 x_1 + w_2 x_2 + c \xrightarrow{g()} \{0, 1\}$$



$$z = w_1 x_1 + w_2 x_2 + w_0$$
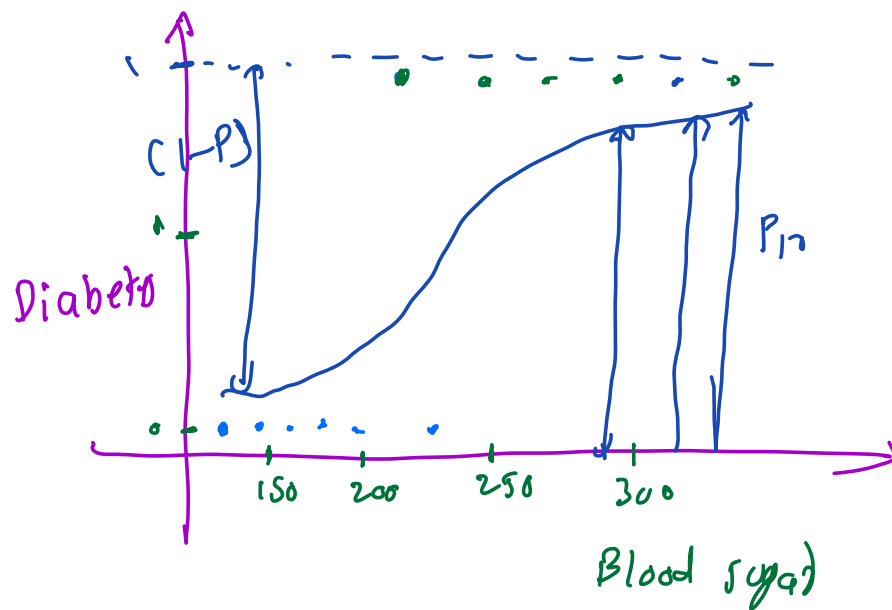
$$\Rightarrow \quad \frac{1}{1 + e^{-z}}$$

| Blood sugar level | Diabetic |
|---|---|
| 190 | No |
| 240 | Yes |
| 100 | Yes |
| 16 | No |
| — | — |
| — | — |



Diabetic

$(1-P)$

$P_n$

Blood sugar

150  200  250  300

$> 0.5 \rightarrow$ yes

$$\leq 0.8 \rightarrow \text{no}$$

$$\frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

$$\Rightarrow (1-P_1) \cdot (1-P_2) \cdot (1-P_3) - \ldots \ldots - (P_7) \cdot$$
$$(P_0) \cdot (P_9)$$

Likehood

$$\Rightarrow \quad P = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

$$1 - P = \frac{e^{-(\beta_0 + \beta_1 x)}}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

$$\frac{P}{1-P} = e^{(\beta_0 + \beta_1 x)} \qquad \frac{P}{1-P}$$

$$\log \left( \frac{P}{1-P} \right) = \beta_0 + \beta_1 x$$

$$P \rightarrow \sim s$$

$$\overline{1-p}$$

$$P = 3(1-p)$$

Train $\Rightarrow$ model $f(x)$

$$\Downarrow$$

$$R^2 \Longleftarrow CV$$

Test $\Rightarrow R'^2$

$$1 - \left( - \right)$$