Data

Train

Train

model fit

Validation

model score

Test

Final Evaluation

(Hold out)

Comparing

1. Model 1 → Valid score
2. Model 2 → valid score
3. Model 3 → valid score

Select the Best Model

→ Hyperparameter Tuning

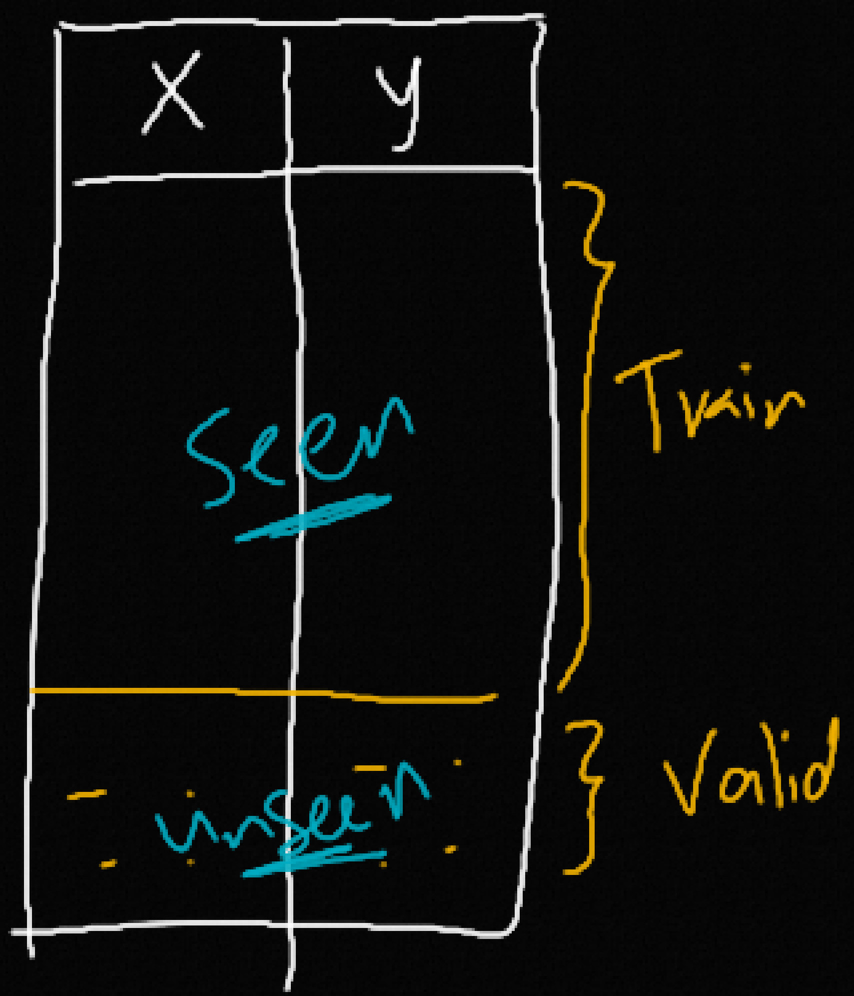
select the Best

* Evaluation on Train data → optimistic
* Selection on Test data → cheating

① Shuffle

② Split
Random

Train Data



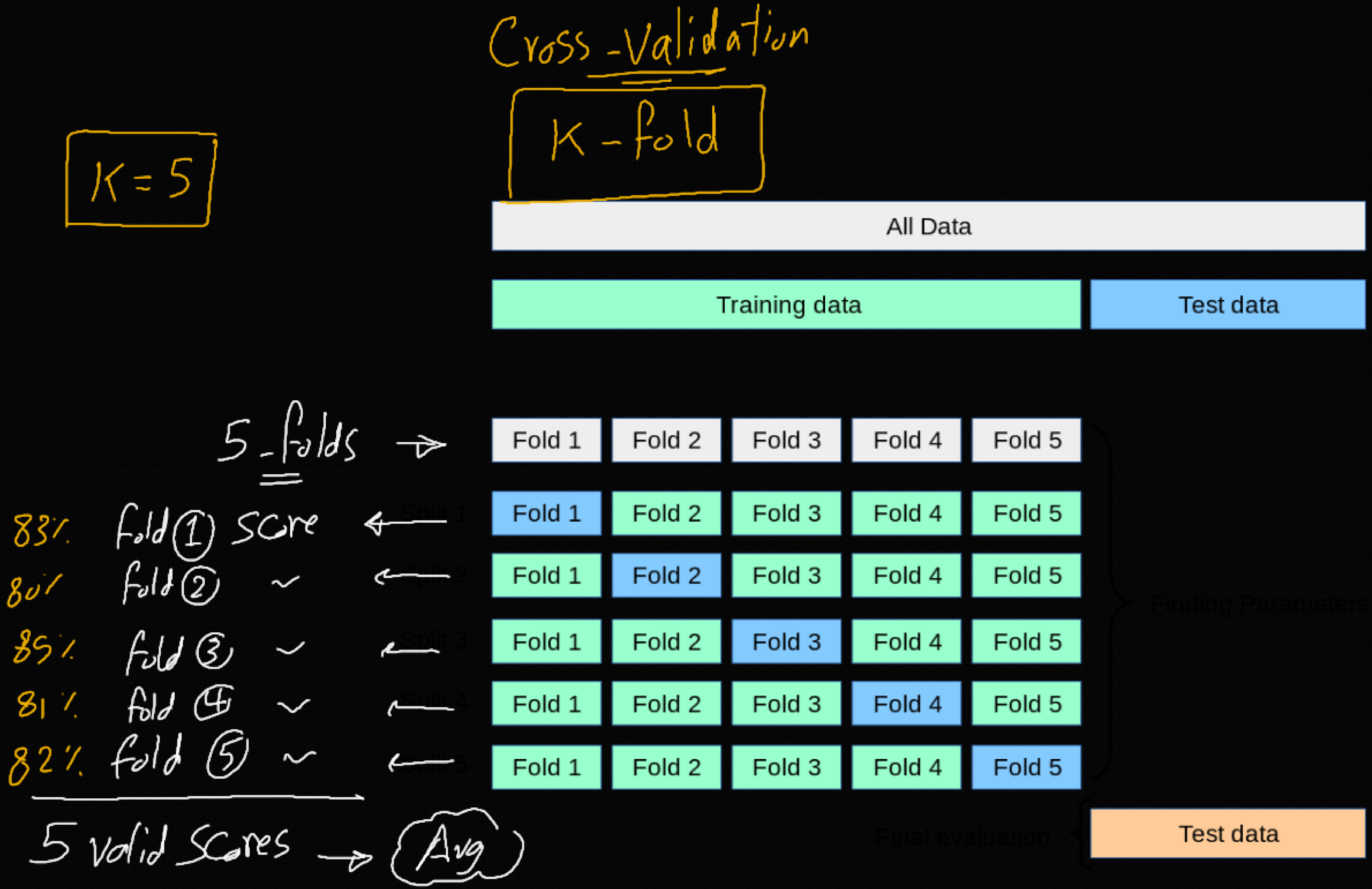
⇒ Ensure Best Training (Seen)

⇒ ~ Best validation (unseen)

⇓

Repeating

Cross_validation


Cross-Validation
K-fold
K=5
All Data
Training data
Test data
5 folds
Fold 1
Fold 2
Fold 3
Fold 4
Fold 5
83% fold 1 score
80% fold 2
85% fold 3
81% fold 4
82% fold 5
5 valid scores → Avg
Test data

Hyperparameter Tuning

Ridge with polynomial

alpha   degree

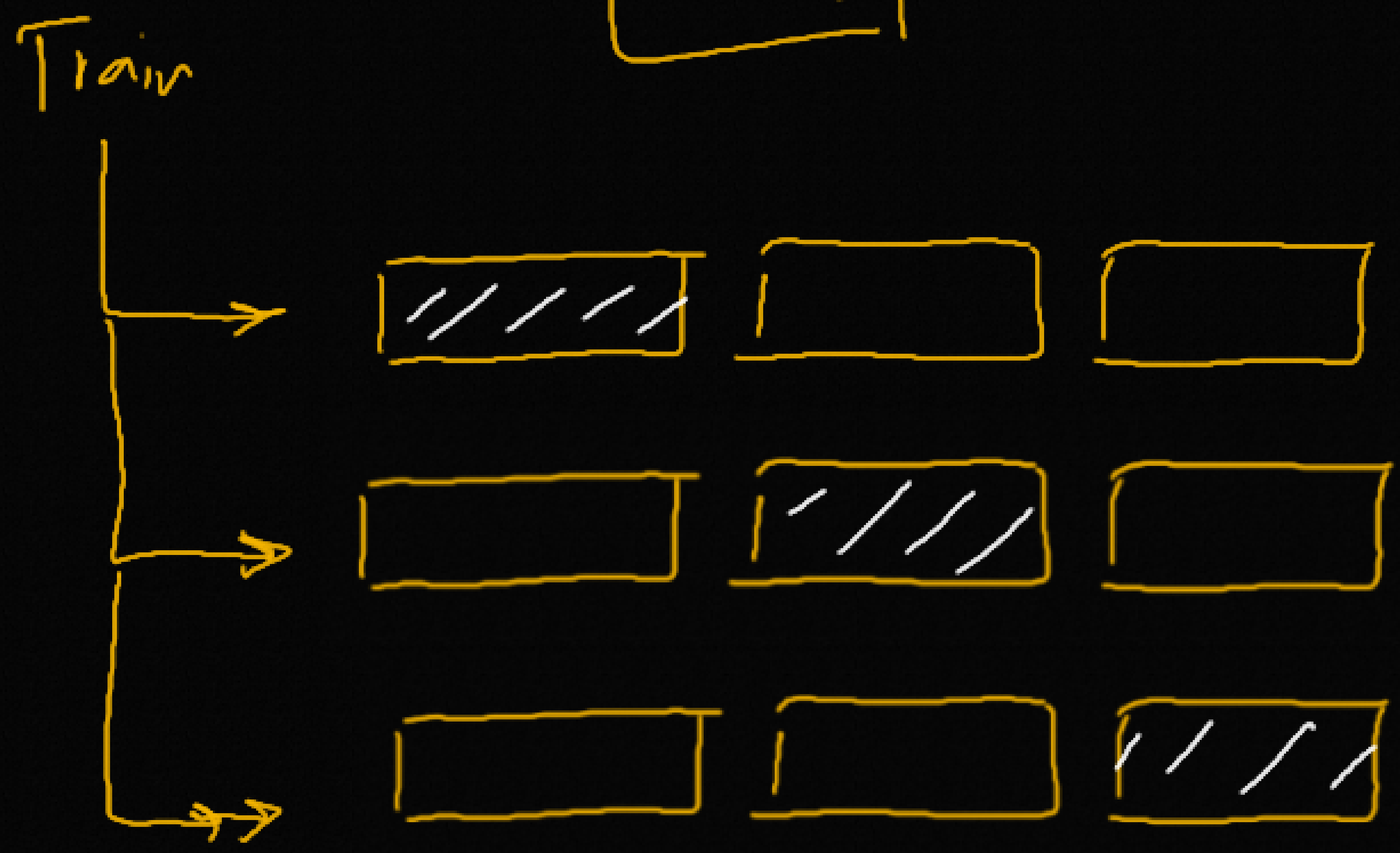
```
Scores = [ ]

for d in [2, 3, 4]:
    for alpha in [0.1, 1, 10]:
        poly(d)
        Ridge(alpha)
        fit
        CV(s) → avg
    scores.append(...)
```

Cross validation

→ Select the hyperparameters with the best score

$\boxed{CV=3}$



Repeat for 9

1. d=2, alph=0.1
2. d=2, ~ = 1
3. d:2, ~ = 10
4. d:3, {
5.
6.
7. d:4 {
8.
9.

Logistic Regression (clf Model)

Linear Model

Linear Eqn → $\boxed{\phi}$ → probability → class 1 / class 0

$p(\text{class } 1)$

f

$[-\infty \to \infty]$ $W^T X$ = Cont. Nr. ... $[0 \to 1]$

$\boxed{v}$ Reg

* Spam filter

Lg Reg

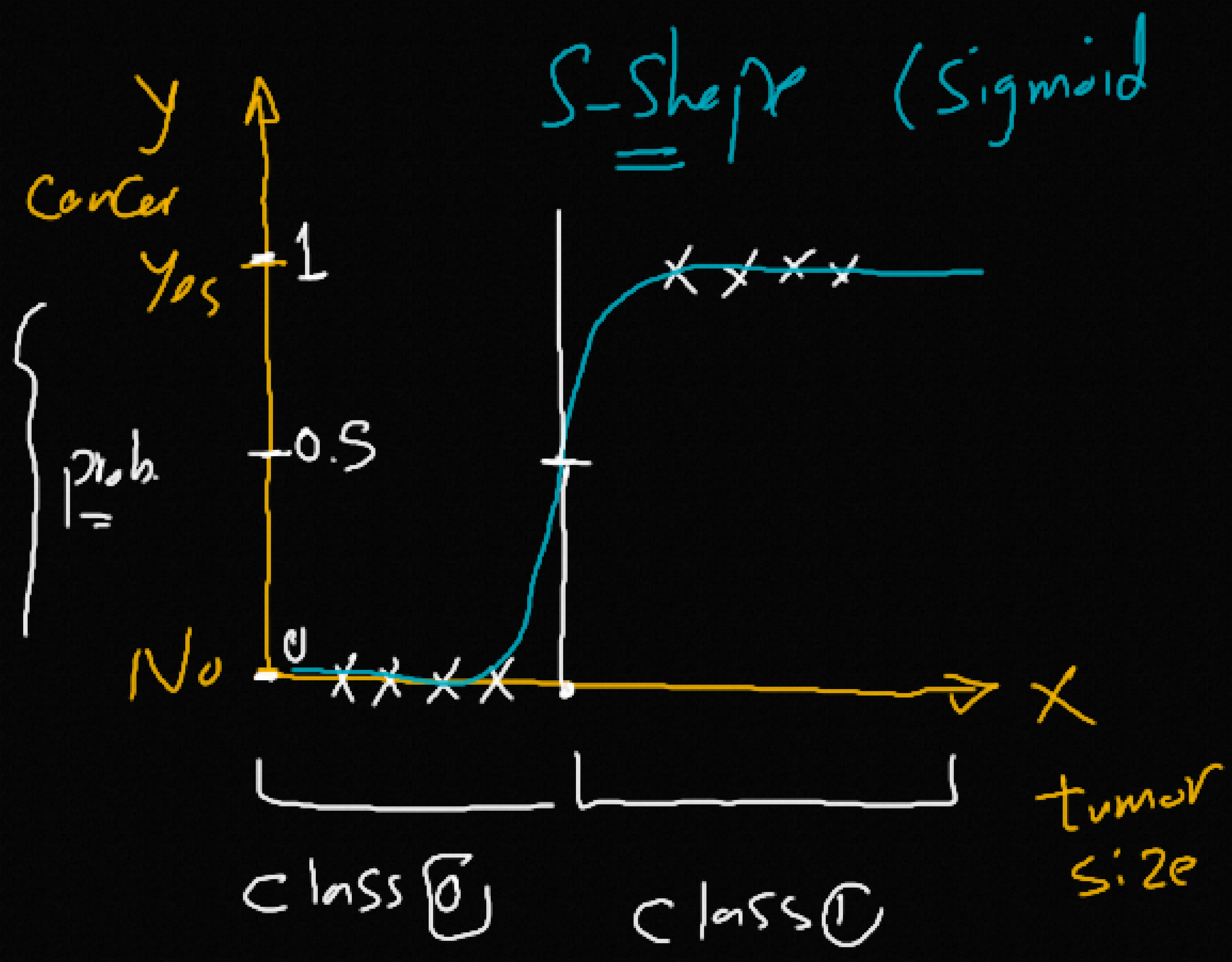
Email Data → $\boxed{\text{Lin Eqn} \to \sigma}$ → Spam / Not

$(-\infty \to +\infty]$ $\qquad$ $(0 \to 1]$

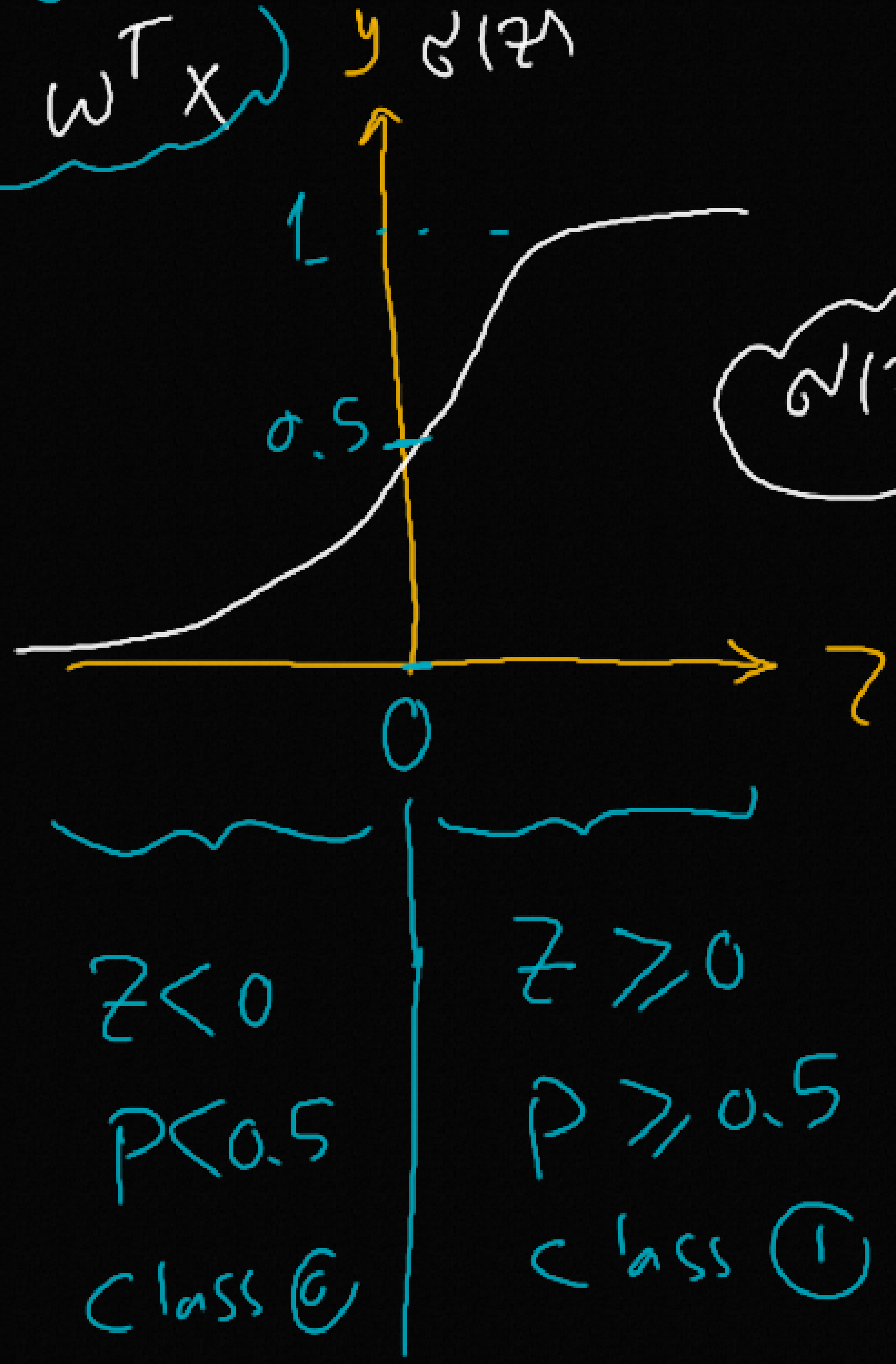
$P(\text{Spam})$

$$\underbrace{0 \;\text{———}\; }_{\text{class } 0}\; 0.5 \;\underbrace{\text{———}\; 1}_{\text{class } 1}$$

(Not spam) class 0 $\qquad$ class 1 (Spam)



$$Z = W^T X$$



$$\sigma(z) = \frac{1}{1+e^{-z}}$$

$$x_1 \quad x_2 \quad x_3 \; \cdots \quad x_n \qquad y$$

$$\vdots \quad \vdots \quad \vdots \qquad \vdots \qquad \vdots$$

$$z = w_0 + w_1 x_1 + w_2 x_2 \cdots + w_n x_n$$ (Can agree po'ly degree)

$$p(y==1) = \sigma(z)$$

$$\boxed{\hat{y} = p(y==1) \geq 0.5}$$

$$\boxed{\hat{y} = w_0 + w_1 x_1 + w_2 x_2 \cdots + w_n x_n \geq 0}$$

* Cost fn (log_loss)
* Gradient Descent

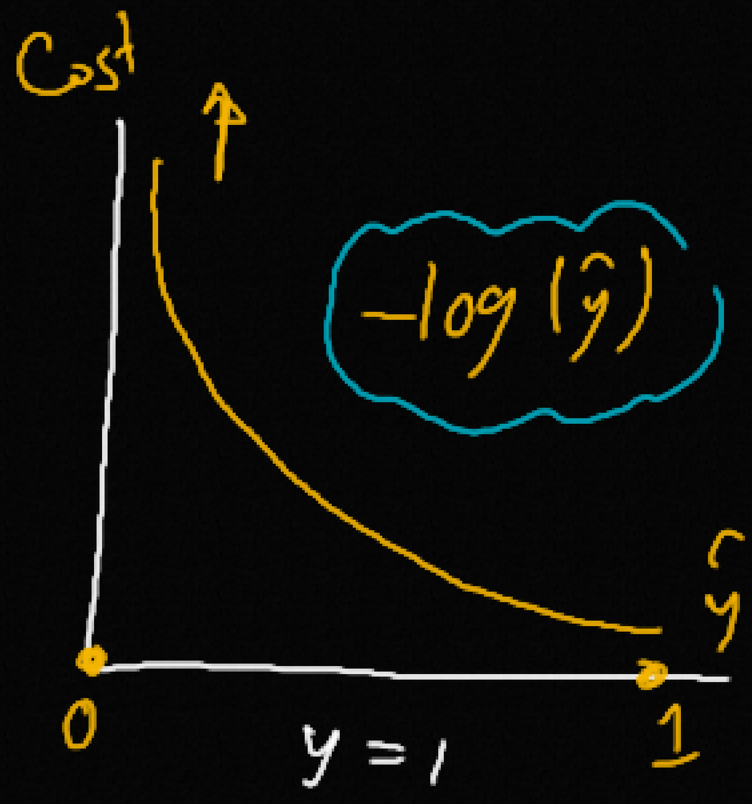
Binay clf

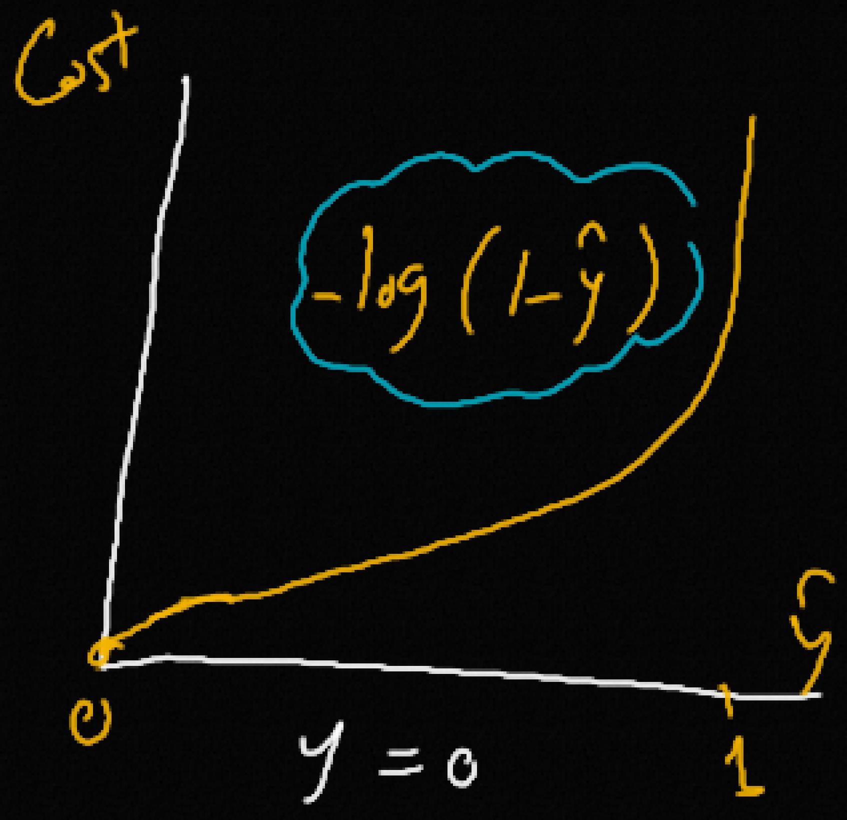
Actual	pred	Cost
0	0	↓
0	1	↑
1	0	↑
1	1	↓

y = 0

y = 1

$$\text{Cost} = \begin{cases} -\log(\hat{y}) & y = 1 \\ -\log(1-\hat{y}) & y = 0 \end{cases}$$





Cost fn

$$\boxed{J = -y\log(\hat{y}) - (1-y)\log(1-\hat{y})}$$

$$J(w) = \frac{-1}{m} \sum_{i=1}^{m} y_i \log(\sigma(w^T x_i)) + (1 - y_i) * \log(1 - \sigma(w^T x_i))$$

Use Grad. Descent to find best $\underline{\underline{W}}$

Lin Reg

Log. Reg

Control of Penalty

alpha $\propto$ penalty

$C \propto \frac{1}{\text{penalty}}$