

Supervised ML

Regression

Linear Regression

- Bias vs. Variance
- Regularization
- Cross Validation
- Hyperparameter Tuning

Error / R² Score

Classification

Logistic Regression

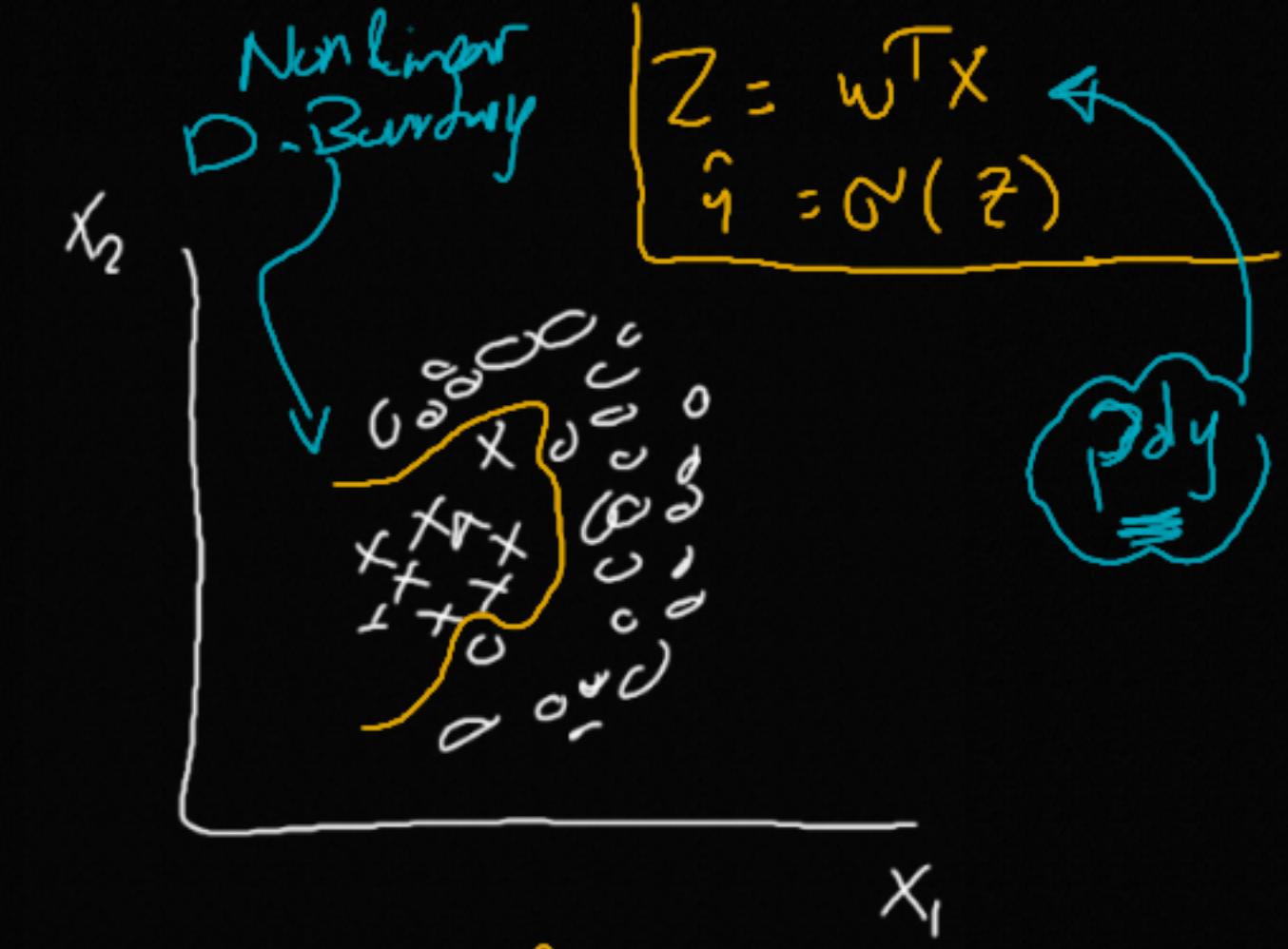
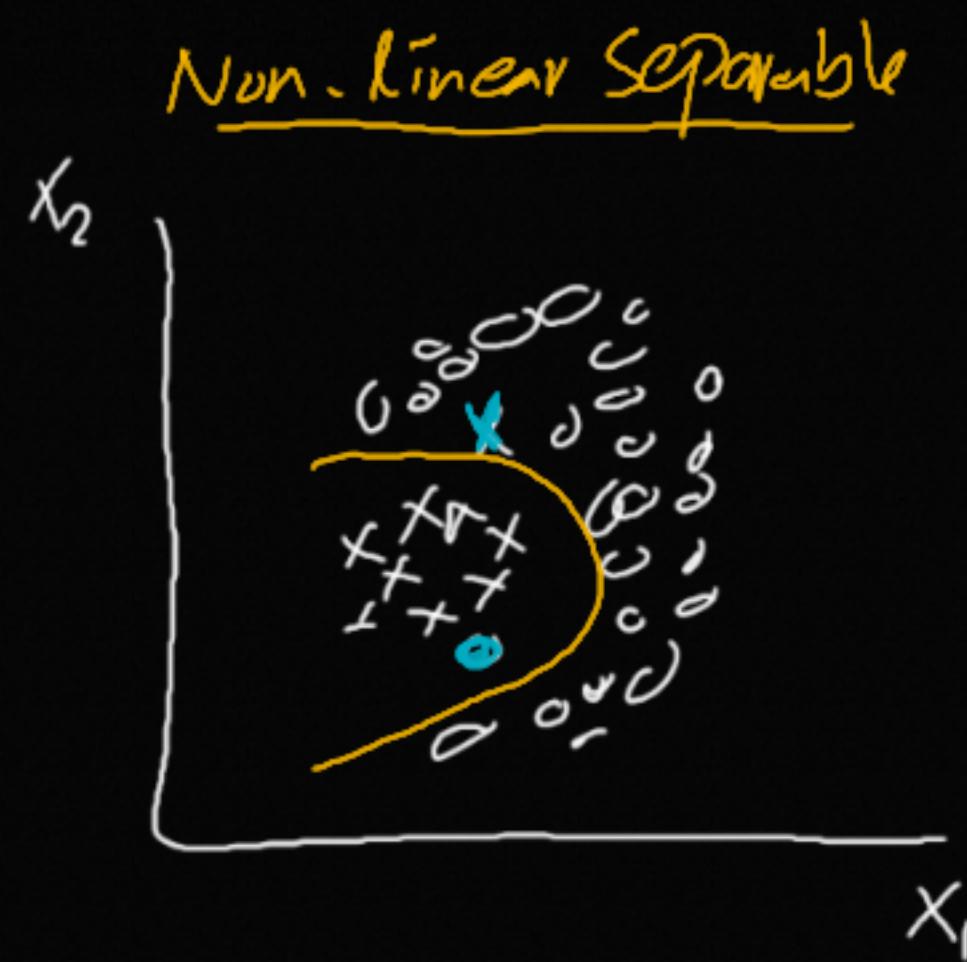
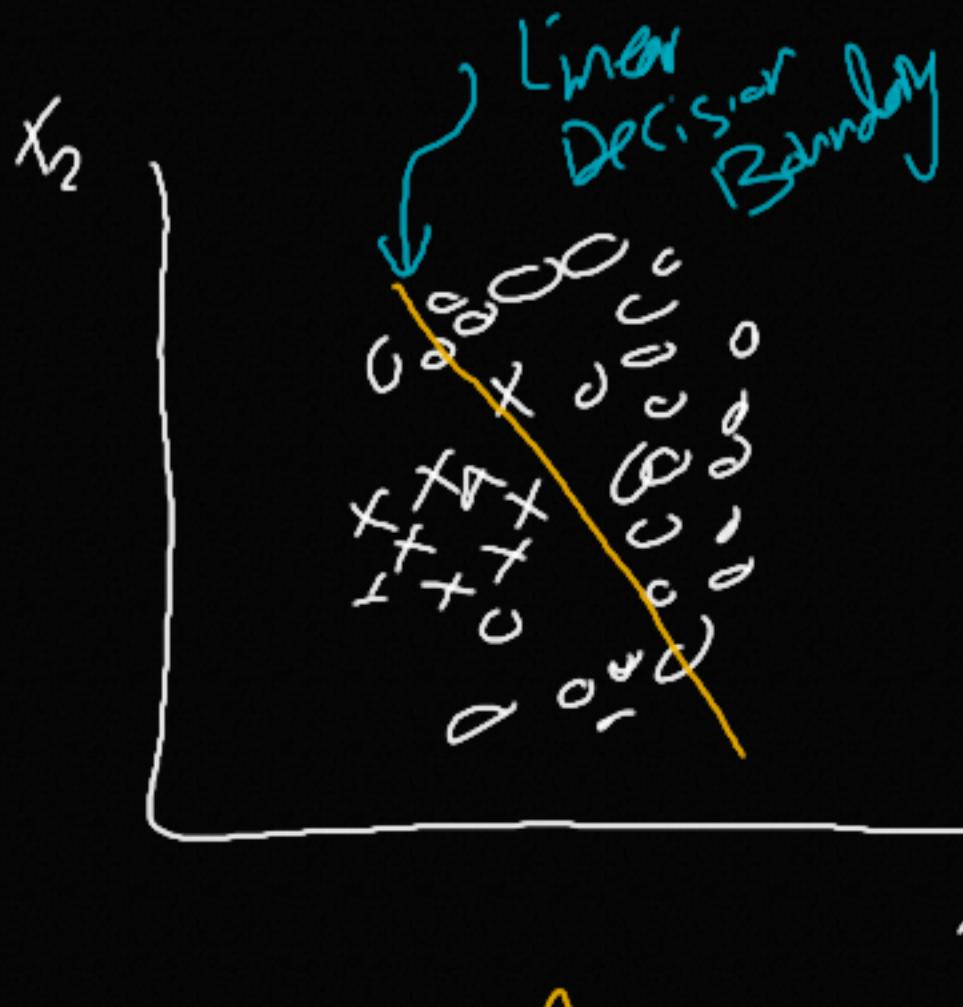
$$(C) \propto \frac{1}{\text{Penalty}}$$

① ✅ Underfitting vs. Overfitting

② ✅ Multi-class classification

③ Evaluation Metrics

Confusion Matrix / Classification Report



Underfitting

Train ACC

Valid ACC

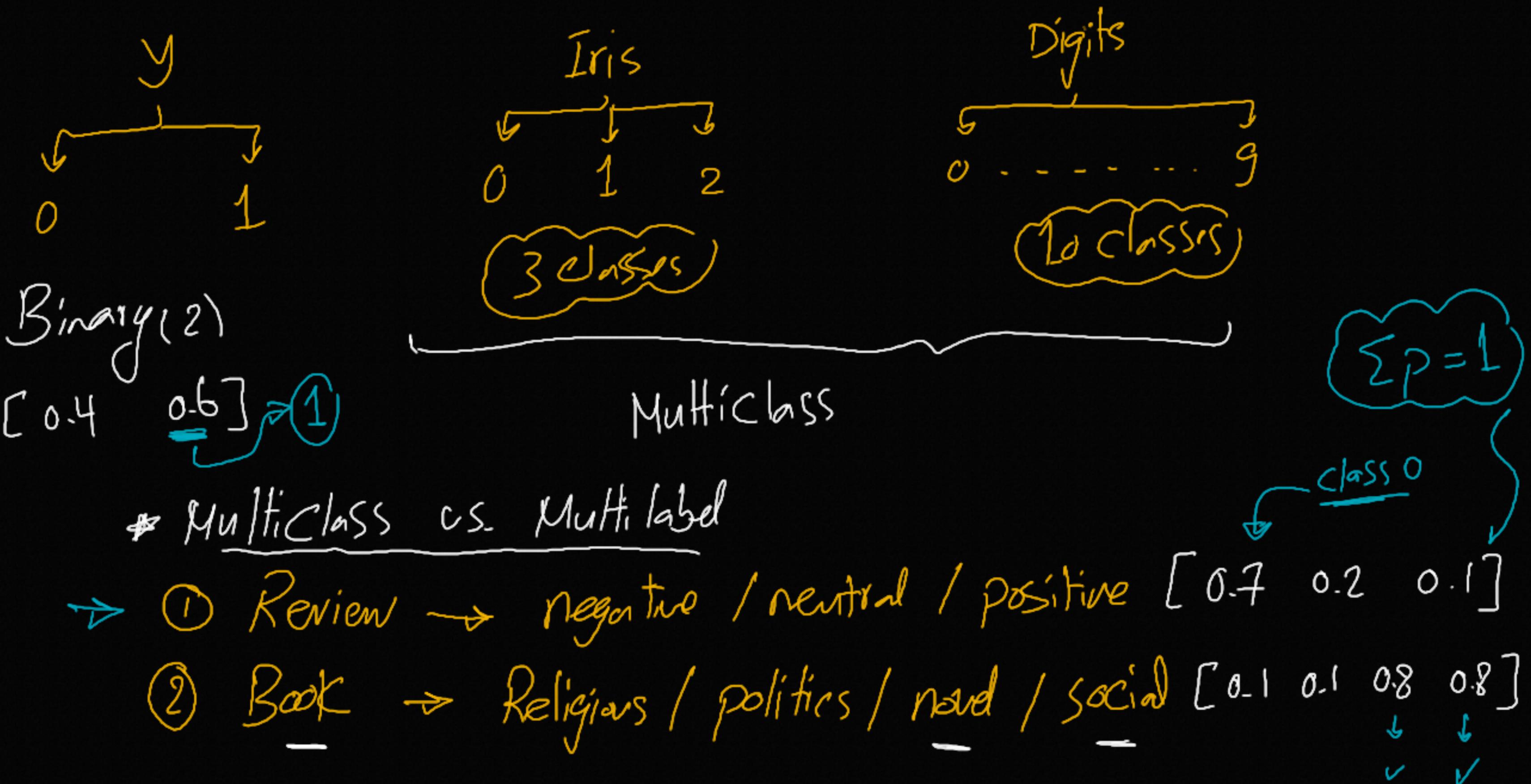
Good fit



Overfitting



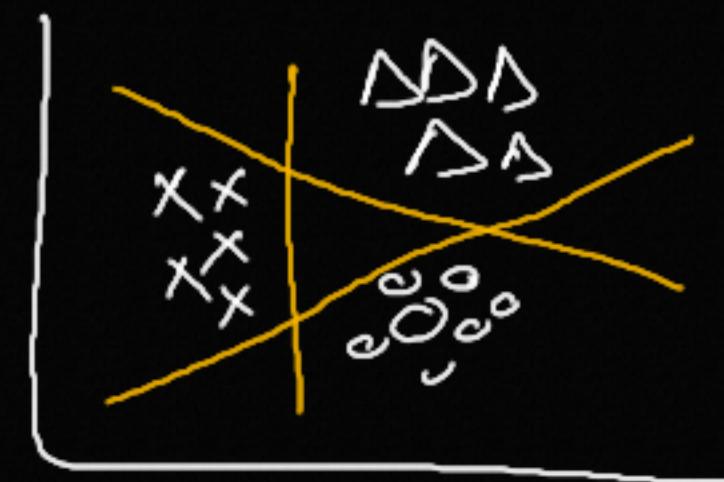
Multiclass classification



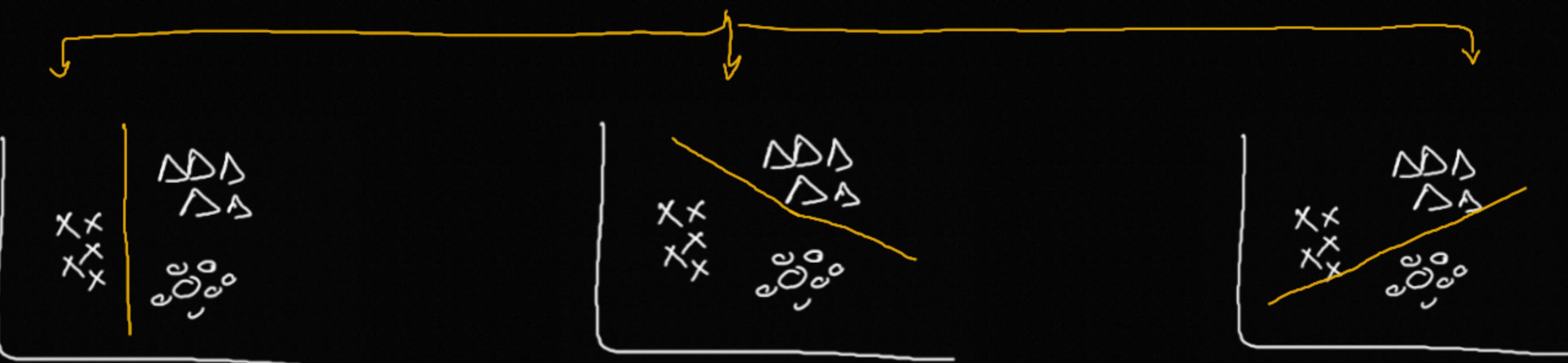
Multi-class

① OVR \rightarrow one-vs-rest

Binary classifiers = # classes



3 classes



② one-vs-one

$$\# \text{ Binary } = \frac{n(n-1)}{2} \quad n = \# \text{ classes}$$

classes [0, 1, 2, 3]

OVR

0 \leftrightarrow All

1 \leftrightarrow All

2 \leftrightarrow All

3 \leftrightarrow All

45. clfs

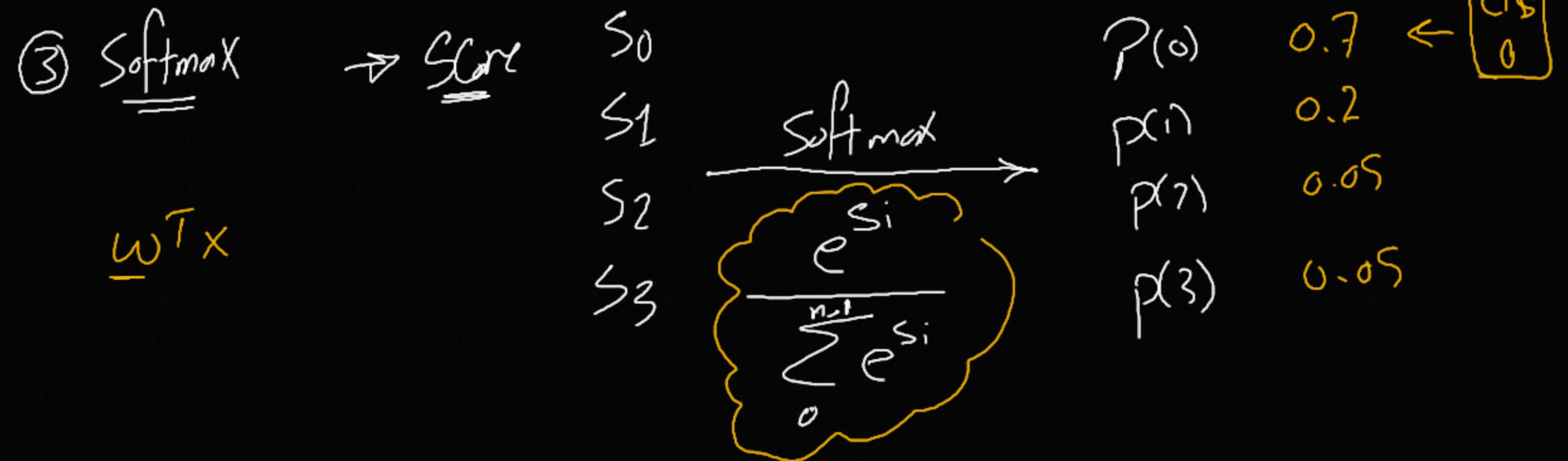
OVO

0 \leftrightarrow 1 / 0 \leftrightarrow 2 / 0 \leftrightarrow 3

1 \leftrightarrow 2 / 1 \leftrightarrow 3

2 \leftrightarrow 3

6 B. clfs



Evaluation Metrics

① Regression Metrics

Errors

diff.
 y_{act} y_{pred}

MSE

MAE

RMSE

=

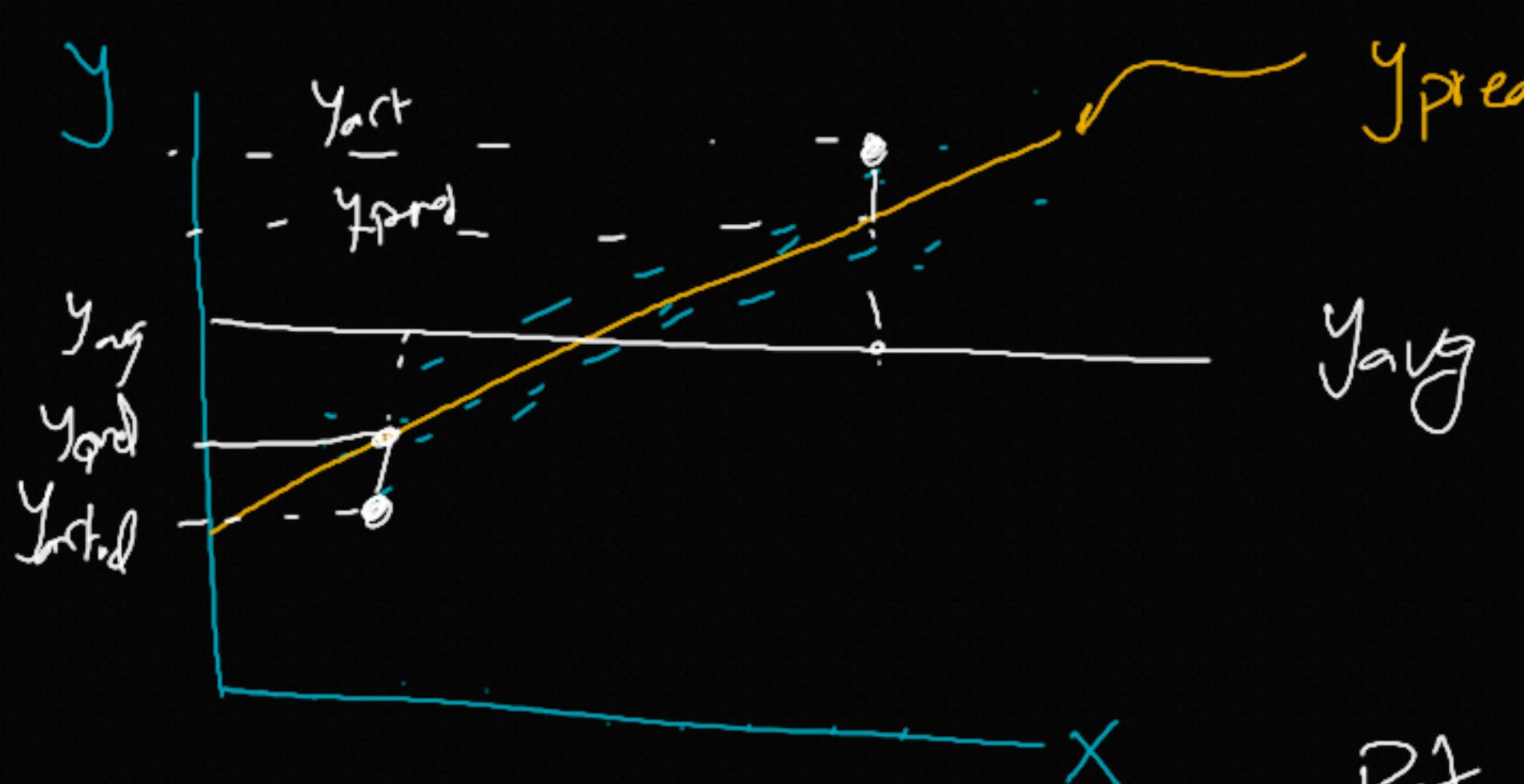
Relative

M1/M2/M3

R2 Score

Ref. \rightarrow y_{avg}

$$R^2 = 1 - \frac{\sum_{i=1}^m (y_{pred} - y_{actual})^2}{\sum_{i=1}^m (y_{avg} - y_{actual})^2}$$



R₂ Score (Regression Acc)

$$R^2 = 1 - \frac{\sum_{i=1}^m (y_{\text{pred}} - y_{\text{actual}})^2}{\sum_{i=1}^m (y_{\text{avg}} - y_{\text{actual}})^2}$$

Best $\Rightarrow y_{\text{pred}} = y_{\text{actual}} \rightarrow R^2 = 1$

Worst $\Rightarrow y_{\text{pred}} = y_{\text{avg}} \rightarrow R^2 = 0$

You Model Err > Dummy Reg Err

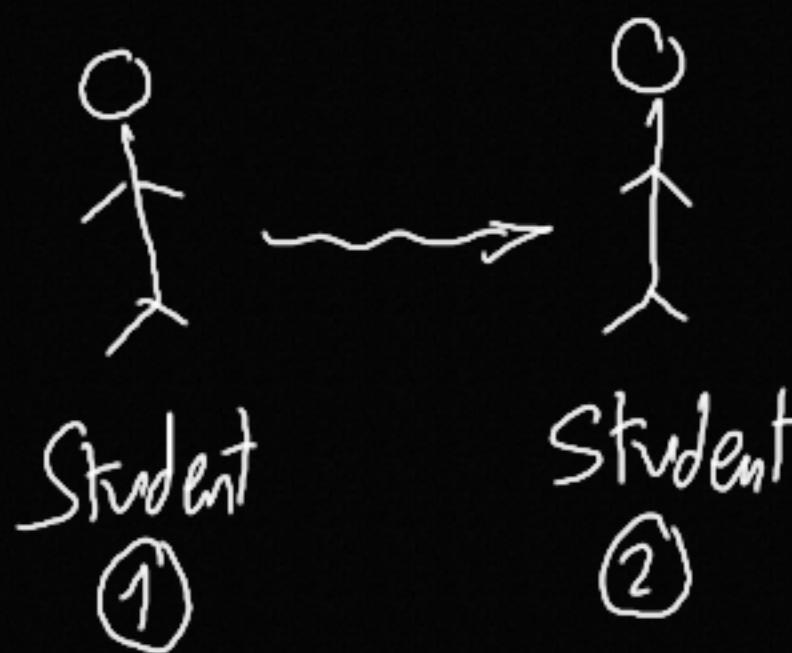
Dummy Reg

\Rightarrow Adjusted R₂ Score

Classification Metrics

① Accuracy

($Y_{pred} == Y_{actual}$)



Imbalance
=

8 [1]
2 [0]

10 [1]

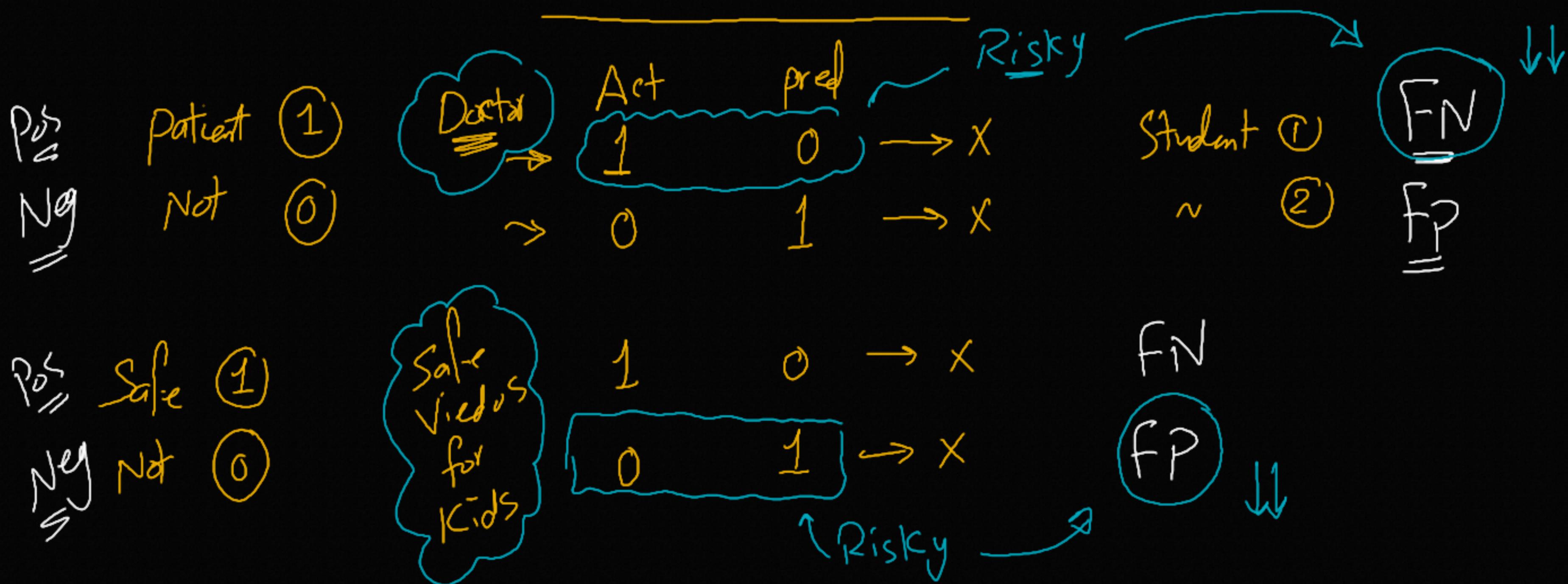
→ 80% Accuracy

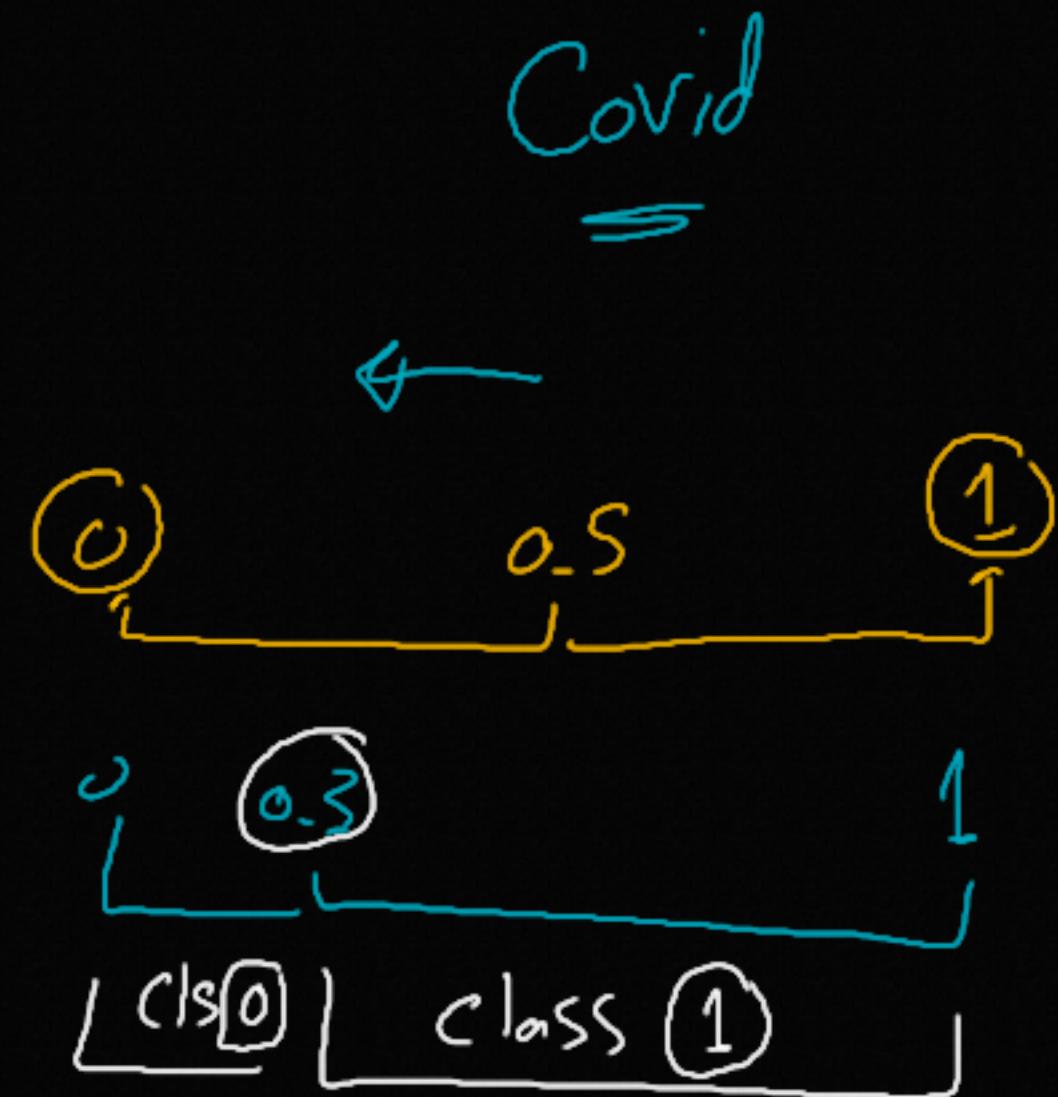
	Yact		Ypred
	0	0	✓
→ 0	0	1	✗
→ 1	1	0	✗
1	1	1	✓

Dummy cl

* Dummy Reg \rightarrow $y_{pred} = y_{avg}$

* Dummy clf \rightarrow $y_{pred} = \underline{y_{mode}}$ (most frequent)





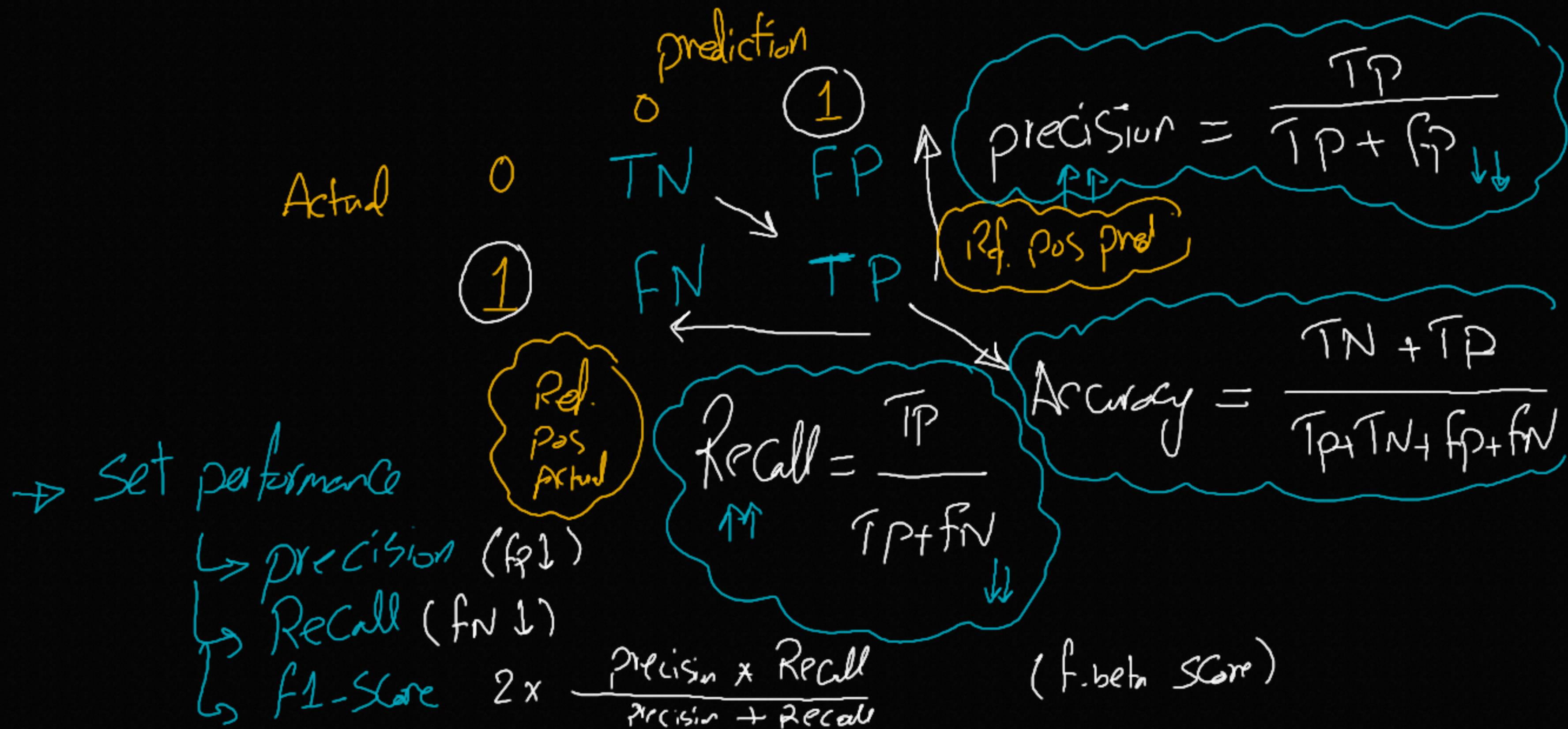
Default
prob.
Threshold
0.5

Safe Videos



Confusion Matrix

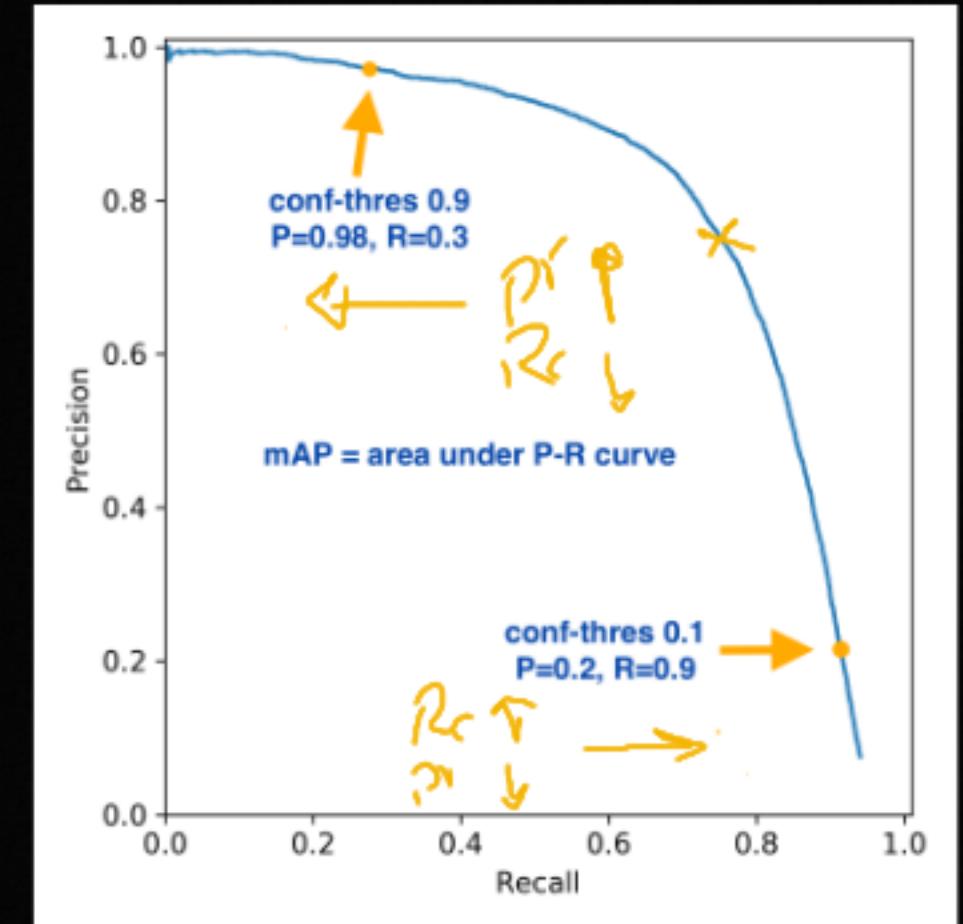
n - classes
shape $(n \times n)$



Classification Report

	precision	recall	f1-score	support
0	0.77	0.86	0.81	37584
1	0.84	0.75	0.79	37577
accuracy			0.80	75161
macro avg	0.81	0.80	0.80	75161
weighted avg	0.81	0.80	0.80	75161

Precision-Recall Curve

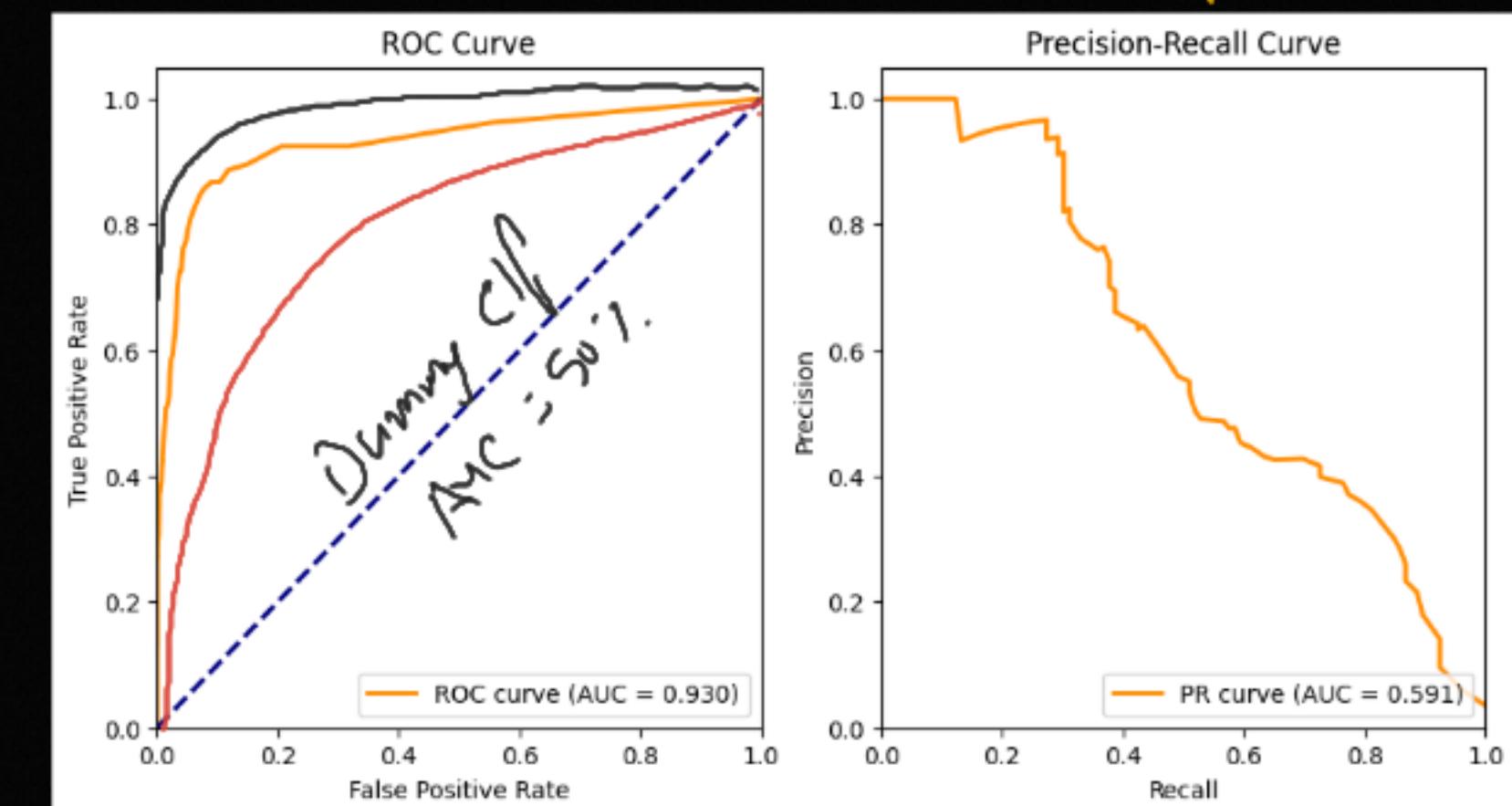


Area under curve

AUC Score

Probability Threshold
0.1 → 0.9

ROC-Curve (TPR, FPR)



R₂ Score (Test)

→ reg_model.Score (Xtest, Ytest)
→ R₂-Score (Ytest, Ypred)

Ypred = Reg_model.predict (Xtest)

clf Accuracy (Test)

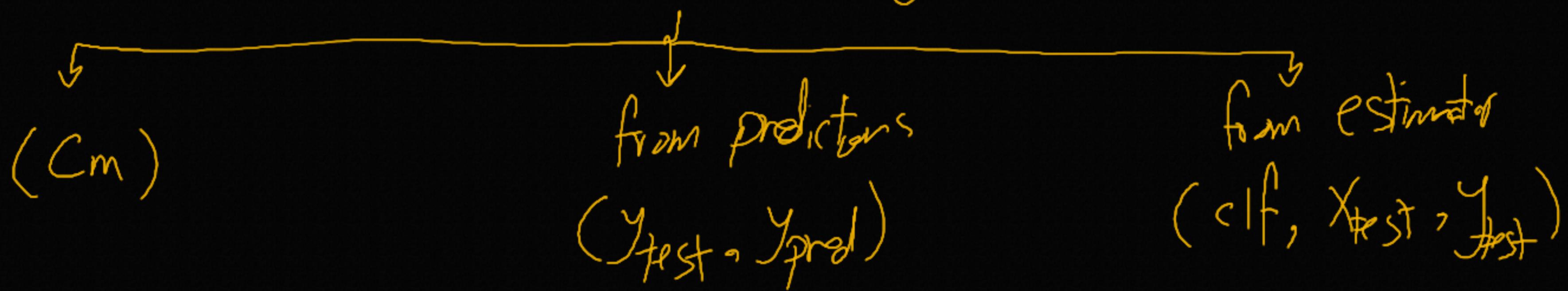
→ $\text{clf.score}(X_{\text{test}}, Y_{\text{test}})$

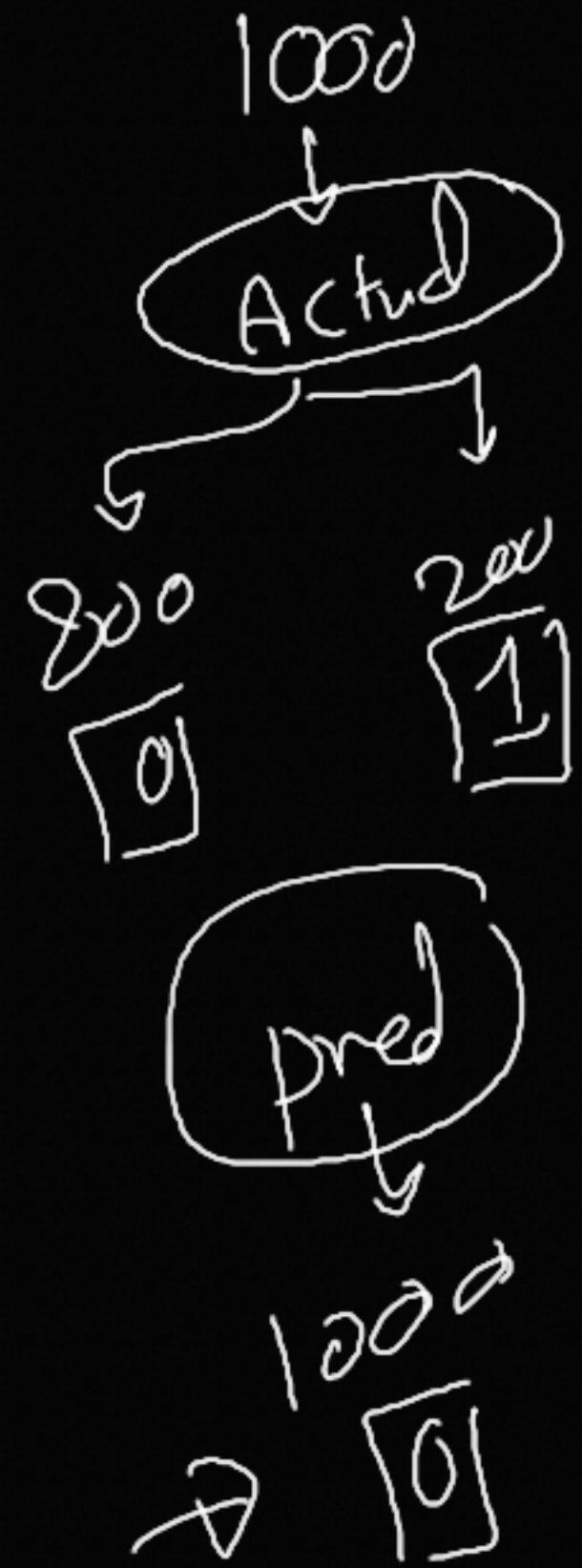
$y_{\text{pred}} = \text{clf.predict}(X_{\text{test}})$

accuracy-Score (Y_{test} , y_{pred})

Confusion Matrix \rightarrow Array $(n \times n)$ $n = \# \text{ classes}$

Confusion Matrix Display





	Pred	
	0	1
0	800	0
1	200	0

Dummy Clf → Model

Hyperparameter Tuning

Param-grid = { — : —
 — : —
 . . .
 . . . }

model ✓

(model, param-grid, Xtrain, ytrain)

Scoring ↗
Reg → 'R2-Score'
clf → 'Accuracy-Score'
'precision'
'recall'
'f1'