

Metrics to Evaluate Model Performance

After this video you will be able to..

- Discuss how performance metrics can be used to evaluate models
- Name three model evaluation metrics
- Explain why accuracy may be misleading

Classification

Is this animal a mammal?

Yes

No



Class Labels

Types of Classification Errors

True Label	Predicted Label	Error Type
Yes	Yes	True Positive (TP)
No	No	True Negative (TN)
No	Yes	False Positive (FP)
Yes	No	False Negative (FN)

Is this animal a mammal?

Yes

No

Class Labels

Accuracy Rate

True
Yes
No
No
Yes

Predicted
Yes
No
Yes
No

Error

True Positive (TP)
True Negative (TN)
False Positive (FP)
False Negative (FN)

$$\begin{aligned}\text{Accuracy Rate} &= \frac{\# \text{ correct predictions}}{\# \text{ total predictions}} \\ &= \frac{TP + TN}{TP + TN + FP + FN}\end{aligned}$$

Error Rate

True	Predicted
Yes	Yes
No	No
No	Yes
Yes	No

Error

- True Positive (TP)
- True Negative (TN)
- False Positive (FP)
- False Negative (FN)

$$\begin{aligned}\text{Error Rate} &= \frac{\text{\# incorrect predictions}}{\text{\# total predictions}} \\ &= \frac{\text{FN} + \text{FP}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \\ &= 1 - \text{Accurate Rate}\end{aligned}$$

True Label	Predicted Label
Yes	No
No	No
No	No
Yes	Yes
Yes	Yes
No	No
Yes	No
Yes	Yes
No	No
No	Yes

True	Predicted	Error
Yes	Yes	True Positive (TP)
No	No	True Negative (TN)
No	Yes	False Positive (FP)
Yes	No	False Negative (FN)

Classification Example

True Label	Predicted Label
Yes	No
No	No
No	No
Yes	Yes
Yes	Yes
No	No
Yes	No
Yes	Yes
No	No
No	Yes

True	Predicted
Yes	Yes
No	No
No	Yes
Yes	No

Error

True Positive (TP)
True Negative (TN)
False Positive (FP)
False Negative (FN)

TP= 3

Classification
Example

True Label	Predicted Label
Yes	No
No	No
No	No
Yes	Yes
Yes	Yes
No	No
Yes	No
Yes	Yes
No	No
No	Yes

True
Yes
No
No
Yes

Predicted
Yes
No
Yes
No

Error

True Positive (TP)
True Negative (TN)
False Positive (FP)
False Negative (FN)

TN = 4

Classification
Example

Accuracy Rate

True	Predicted	Error
Yes	Yes	True Positive (TP)
No	No	True Negative (TN)
No	Yes	False Positive (FP)
Yes	No	False Negative (FN)

$$\begin{aligned}\text{Accuracy Rate} &= \frac{\# \text{ correct predictions}}{\# \text{ total predictions}} \\ &= \frac{TP + TN}{TP + TN + FP + FN} \\ &= (3 + 4) / 10 = 7 / 10 = 0.7\end{aligned}$$

Error Rate

True
Yes
No
No
Yes

Predicted
Yes
No
Yes
No

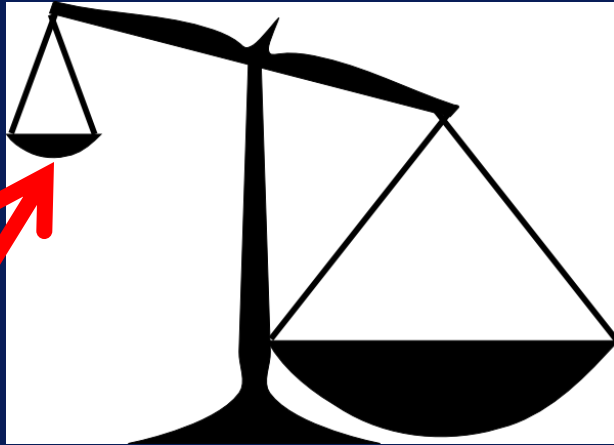
Error

True Positive (TP)
True Negative (TN)
False Positive (FP)
False Negative (FN)

$$\begin{aligned}\text{Error Rate} &= \frac{\# \text{ incorrect predictions}}{\# \text{ total predictions}} \\ &= 1 - \text{Accuracy Rate} \\ &= 1 - 0.7 = 0.3\end{aligned}$$

Limitation with Accuracy

Is this tumor cancerous?



most are
negative
examples

very few
positive
examples

Class Imbalance
Problem

Limitation with Accuracy

Is this tumor cancerous?

- Say 3% of samples are cancer
- If model always predicts non-cancer
 - Accuracy = 97%
 - But no cancer cases detected!

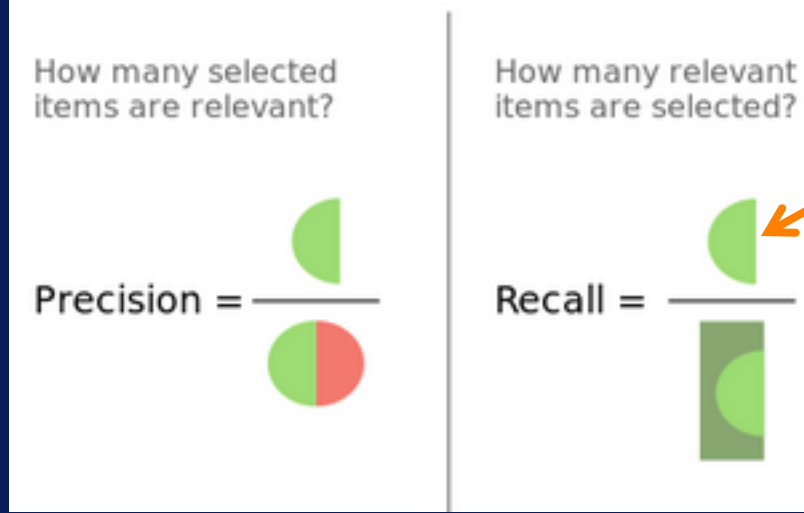
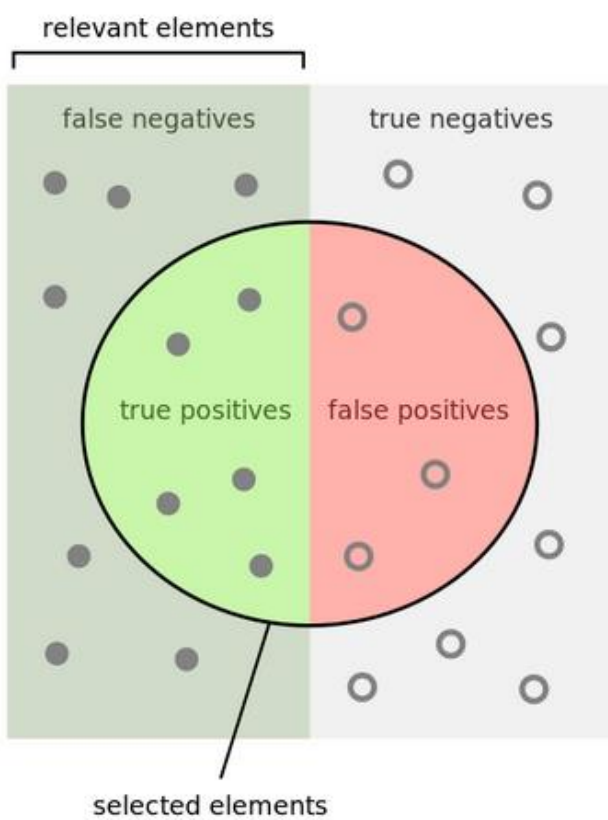


Precision & Recall

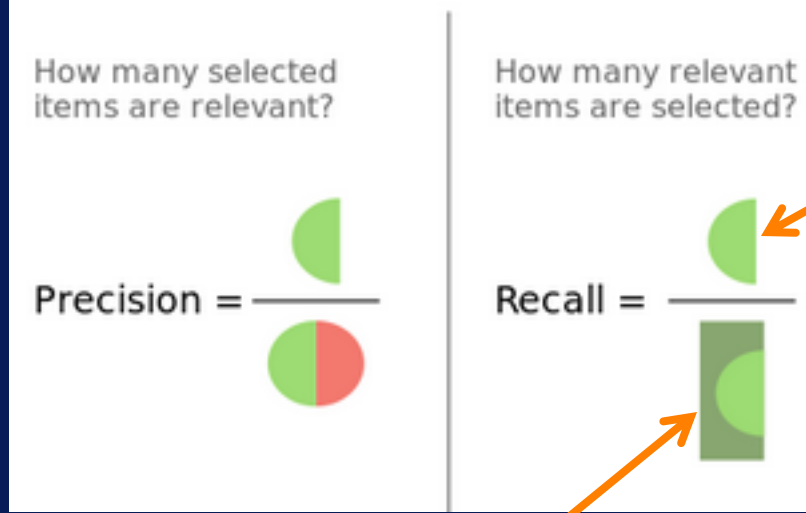
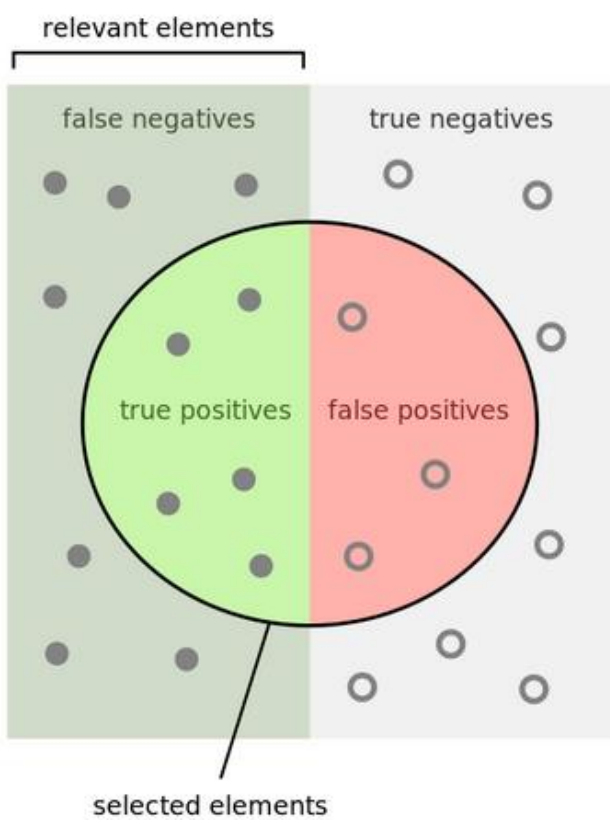
True	Predicted	Error
Yes	Yes	True Positive (TP)
No	No	True Negative (TN)
No	Yes	False Positive (FP)
Yes	No	False Negative (FN)

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \leftarrow \text{All samples with Predicted} = \text{Yes}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \leftarrow \text{All samples with True} = \text{Yes}$$



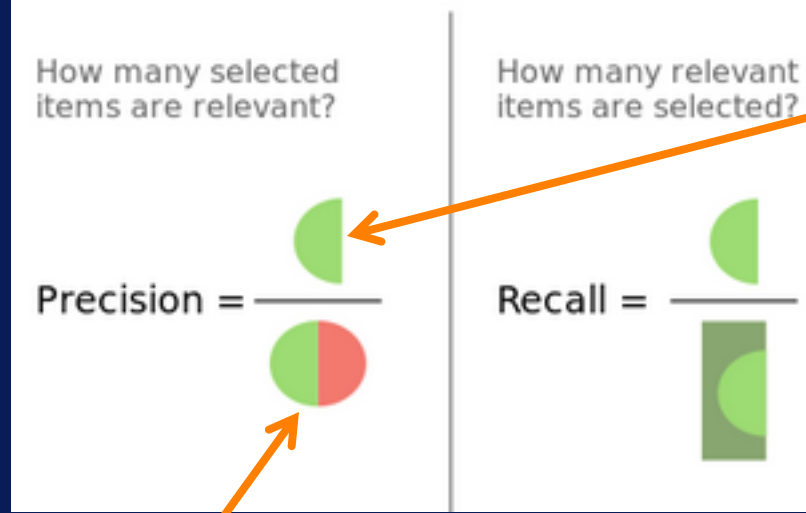
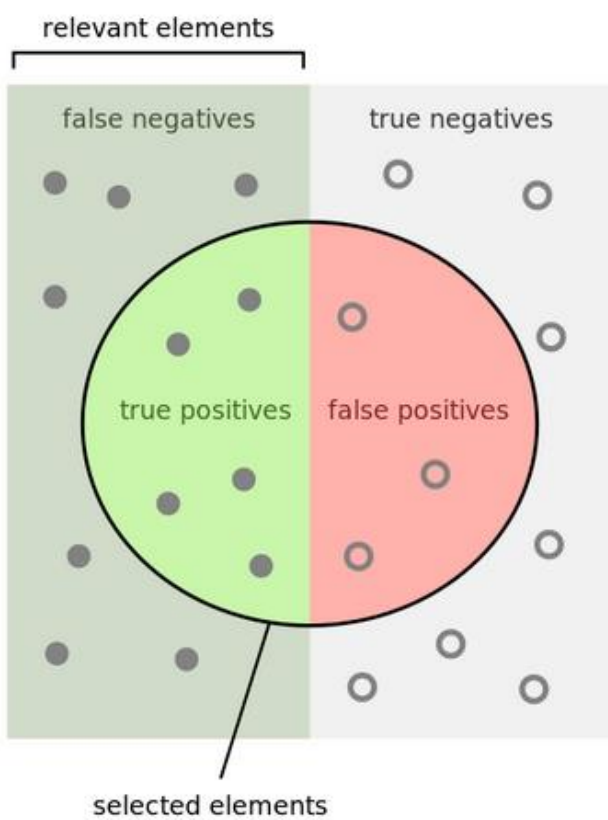
Recall



Samples correctly predicted as Positive

Samples actually Positive

Recall



Samples
correctly
predicted
as Positive

Samples
predicted
as Positive

Recall

Precision & Recall

True	Predicted	Error
Yes	Yes	True Positive (TP)
No	No	True Negative (TN)
No	Yes	False Positive (FP)
Yes	No	False Negative (FN)

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{\text{Positive samples correctly predicted}}{\text{All samples predicted as Positive}}$$

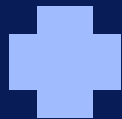
Measure of
exactness

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{\text{Positive samples correctly predicted}}{\text{All samples with true label Positive}}$$

Measure of
completeness

Precision & Recall

Precision



Recall

- Use together
- Goal: Maximize both

F-Measure

Precision



Recall

$$F_1 = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

- F_1 : evenly weighted
- F_2 : weights Recall more
- $F_{0.5}$: weights Precision more

Evaluation Metrics

True	Predicted	Error
Yes	Yes	True Positive (TP)
No	No	True Negative (TN)
No	Yes	False Positive (FP)
Yes	No	False Negative (FN)

*Accuracy
Rate*

*Error
Rate*

*Precision
& Recall*

F_1 -Measure