DS
boosting your data exploration

Evaluate ML System

Ha Noi, 11/12/2022,
kan pham

DS
boosting your data exploration

Content:

1. Accuracy
2. Sensitivity, Specificity and prevalence
3. PPV and NPV
4. Confusion matrix
5. Type error I & II
6. Precision, Recall, and F1-score
7. ROC curve and AUC
8. evaluate multi-classification model

1. Accuracy

DS
boosting your data exploration

How good a classification model is?

$$\text{Accuracy} = \frac{\textit{Examples of correctly of classified}}{\textit{Total number of examples}}$$

1. Accuracy

How good a classification model is?

Positive: *Disease*
Negative: *Normal*

$$\text{Accuracy} = \frac{\textit{Examples of correctly of classified}}{\textit{Total number of examples}}$$

Prediction		Ground Truth: Positive	Ground Truth: Negative
	Positive	20	0
	Negative	20	60

How many Accuracy ?

1. Accuracy

DS boosting your data exploration

Accuracy in Terms of Conditional Probability

Positive: *Disease*
Negative: *Normal*

Accuracy = P(Correct)
= P(Correct, Disease) + P(Correct, Normal)
= ?

		Ground Truth	
		Positive	Negative
Prediction	Positive	20	0
	Negative	20	60

1. Accuracy

DS boosting your data exploration

Accuracy in Terms of Conditional Probability

Positive: *Disease*
Negative: *Normal*

Accuracy = P(Correct)
= P(Correct, Disease) + P(Correct, Normal)
= P(Correct | Disease) P(Disease) + P(Correct | Normal) P(Normal)
= P(+ | Disease) P(Disease) + P(- | Normal) P(Normal)

Sensitivity (TPR) Specificity (TNR)

	Ground Truth	
Prediction	Positive	Negative
Positive	20	0
Negative	20	60

2. Sensitivity and Specificity

DS boosting your data exploration

Accuracy in Terms of Conditional Probability

Positive: *Disease*
Negative: *Normal*

Accuracy = P(Correct)
= P(Correct, Disease) + P(Correct, Normal)
= P(Correct | Disease) P(Disease) + P(Correct | Normal) P(Normal)
= P(+ | Disease) P(Disease) + P(- | Normal) P(Normal)

Sensitivity (TPR) Specificity (TNR)

	Ground Truth	
Prediction	Positive	Negative
Positive	20	0
Negative	20	60

2. Sensitivity and Specificity

DS
boosting your data exploration

Accuracy in Terms of Conditional Probability

Positive: *Disease*
Negative: *Normal*

Accuracy = P(Correct)
= P(Correct, Disease) + P(Correct, Normal)
= P(Correct | Disease) P(Disease) + P(Correct | Normal) P(Normal)
= P(+ | Disease) P(Disease) + P(- | Normal) P(Normal)

Sensitivity (TPR) Specificity (TNR)

*How could you literally define **Sensitivity** and **Specificity** ?*

2. Sensitivity and Specificity

DS boosting your data exploration

Accuracy in Terms of Conditional Probability

Positive: *Disease*
Negative: *Normal*

Accuracy = P(Correct)
= P(Correct, Disease) + P(Correct, Normal)
= P(Correct | Disease) P(Disease) + P(Correct | Normal) P(Normal)
= P(+ | Disease) P(Disease) + P(- | Normal) P(Normal)

Sensitivity (TPR) Specificity (TNR)

*How could you literally define **Sensitivity** and **Specificity** ?*

P(+ | Disease)

If a patient has disease, what is probability that model predict disease?

Sensitivity

P(- | Normal)

If a patient is normal, what is probability that model predicts normal?

Specificity

2. Sensitivity, Specificity and prevalence

DS boosting your data exploration

Accuracy in Terms of Conditional Probability

Positive: *Disease*
Negative: *Normal*

Accuracy = P(Correct)
= P(Correct, Disease) + P(Correct, Normal)
= P(Correct | Disease) P(Disease) + P(Correct | Normal) P(Normal)
= P(+ | Disease) P(Disease) + P(- | Normal) P(Normal)

Sensitivity (TPR) Specificity (TNR)

= Sensitivity * P(Disease) + Specificity * P(Normal)

3. PPV and NPV

P(+ | Disease) → **P(Disease | +)**

Positive: *Disease*
Negative: *Normal*

If a patient has disease, what is probability that model predict disease?

Sensitivity

If a model prediction is positive, what is probability that patient has the disease?

PPV (positive predictive value)

		Ground Truth	
		Positive	Negative
Prediction	Positive	20	0
	Negative	20	60

3. PPV and NPV

P(- | Normal) → **P(Normal | -)**

Positive: *Disease*
Negative: *Normal*

If a patient is normal, what is probability that model predict normal?

Specificity

If a model prediction is negative, what is probability that patient is normal?

NPV (negative predictive value)

	Ground Truth	
Prediction	Positive	Negative
Positive	20	0
Negative	20	60

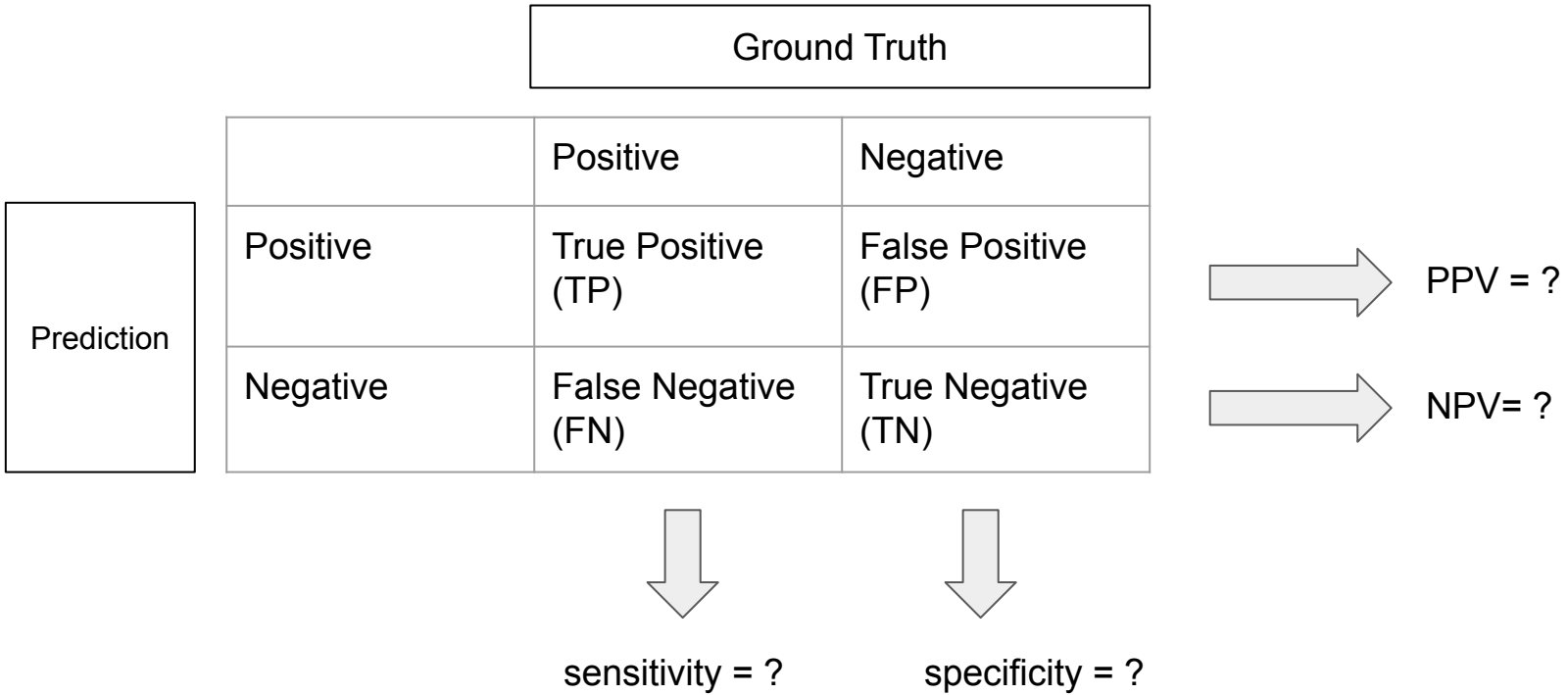
3. PPV and NPV

DS
boosting your data exploration

P(+ | Disease) ***Sensitivity***	**P(Disease | +)** ***PPV***
P(- | Normal) ***Specificity***	**P(Normal | -)** ***NPV***

4. Confusion matrix

DS
boosting your data exploration



4. Confusion matrix

DS boosting your data exploration

		Ground Truth: Positive	Ground Truth: Negative
Prediction	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

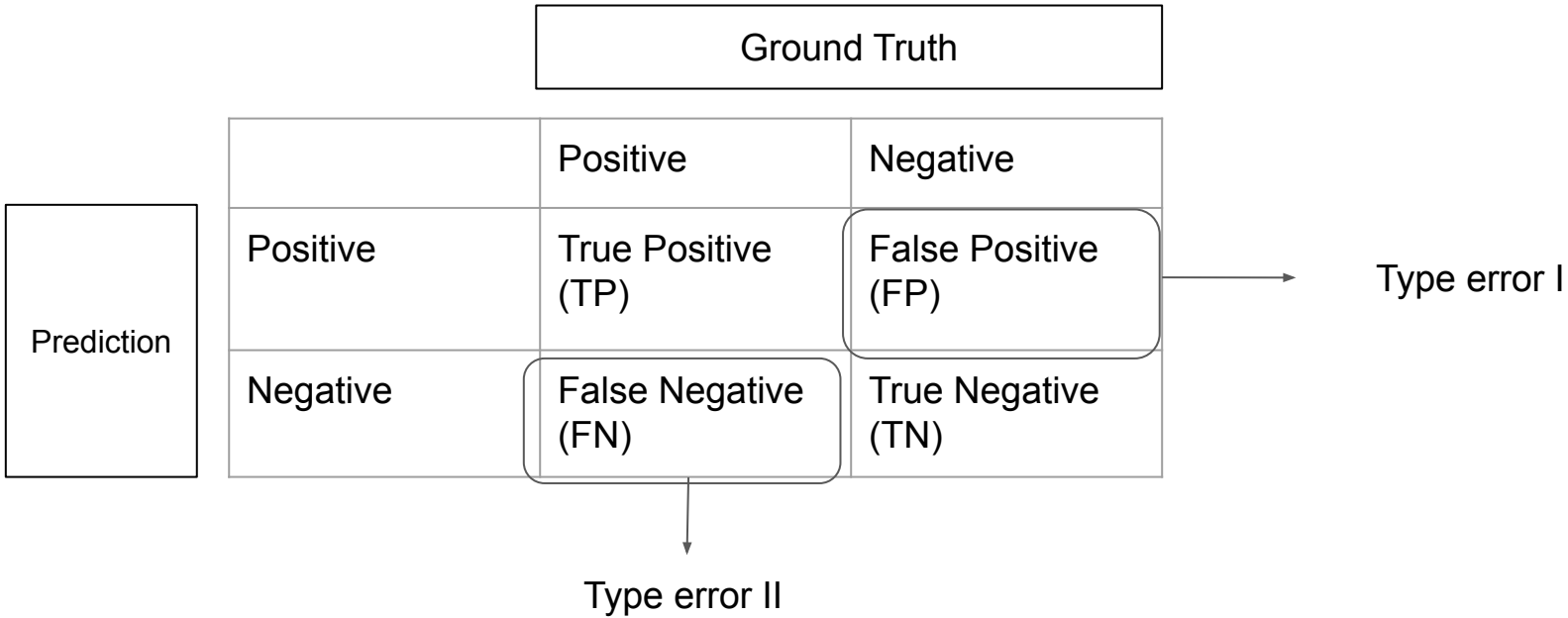
$$\text{PPV} = \frac{TP}{TP + FP}$$

$$\text{NPV} = \frac{TN}{FN + TN}$$

$$\text{sensitivity} = \frac{TP}{TP + FN}$$

$$\text{specificity} = \frac{TN}{FP + TN}$$

5. Type error I & II



What error is more serious?

6. f1-score, precision and recall

DS boosting your data exploration

Model 1 and model 2 is binary classification model have the same accuracy is 80%.

Positive: *Disease*
Negative: *Normal*

Model 1:

		Ground Truth	
Prediction		Positive	Negative
	Positive	0	0
	Negative	20	80

Model 2:

		Ground Truth	
Prediction		Positive	Negative
	Positive	20	0
	Negative	20	60

What is the better model?

6. f1-score, precision and recall

DS boosting your data exploration

Model 1 and model 2 is binary classification model have the same accuracy is 80%.

Positive: *Disease*
Negative: *Normal*

Model 1:

Prediction		Ground Truth: Positive	Ground Truth: Negative
	Positive	0	0
	Negative	20	80

Model 2:

Prediction		Ground Truth: Positive	Ground Truth: Negative
	Positive	20	0
	Negative	20	60

What is the better model?
⇒ Model 2 is better because it predict right 50% patients that get diseases.

6. f1-score, precision and recall

Model 1 and model 2 is binary classification model have the same accuracy is 80%.

Positive: *Disease*
Negative: *Normal*

Model 1:

Prediction		Ground Truth	
		Positive	Negative
	Positive	0	0
	Negative	20	80

Model 2:

Prediction		Ground Truth	
		Positive	Negative
	Positive	20	0
	Negative	20	60

What is the better model?
⇒ Model 2 is better because it predict right 50% patients that get diseases.
Thus, we need to change accuracy metric to another metric.

6. f1-score, precision and recall

DS boosting your data exploration

		Ground Truth: Positive	Ground Truth: Negative
Prediction	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{f1-score} = \frac{2}{1/\text{precision} + 1/\text{recall}}$$

6. f1-score, precision and recall

DS boosting your data exploration

Model 1 and model 2 is binary classification model have the same accuracy is 80%.

Positive: *Disease*
Negative: *Normal*

Model 1:

Prediction		Ground Truth: Positive	Ground Truth: Negative
	Positive	0	0
	Negative	20	80

Precision = ?
Recall = ?
f1-score = ?

Model 2:

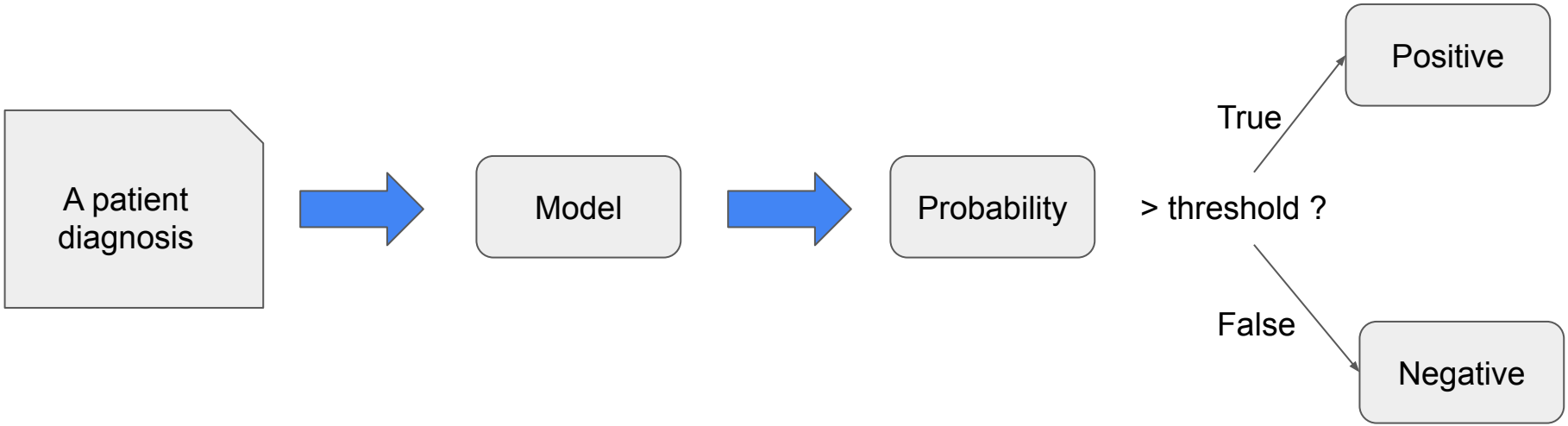
Prediction		Ground Truth: Positive	Ground Truth: Negative
	Positive	20	0
	Negative	20	60

Precision = ?
Recall = ?
f1-score = ?

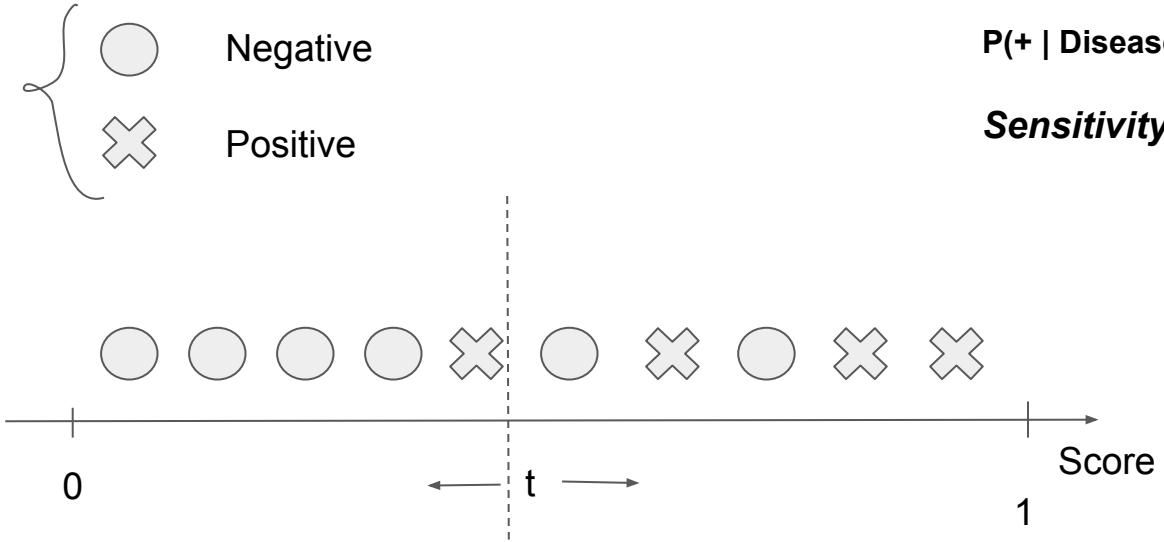
7. ROC curve and threshold

DS boosting your data exploration

What is threshold ?

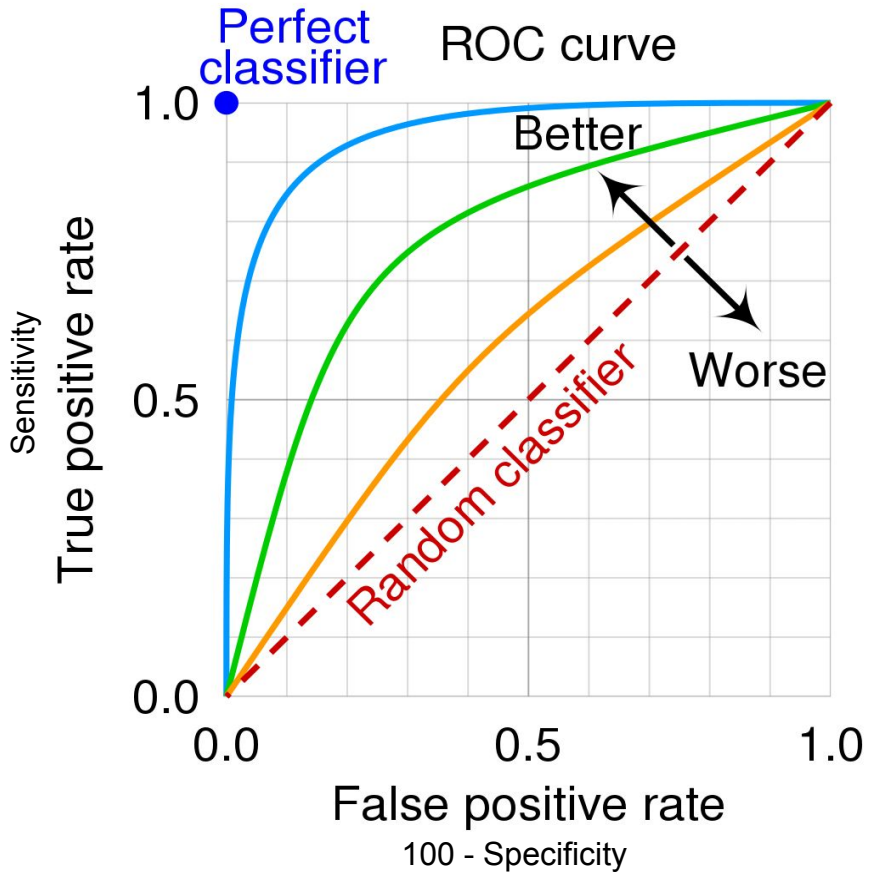


7. ROC curve and threshold



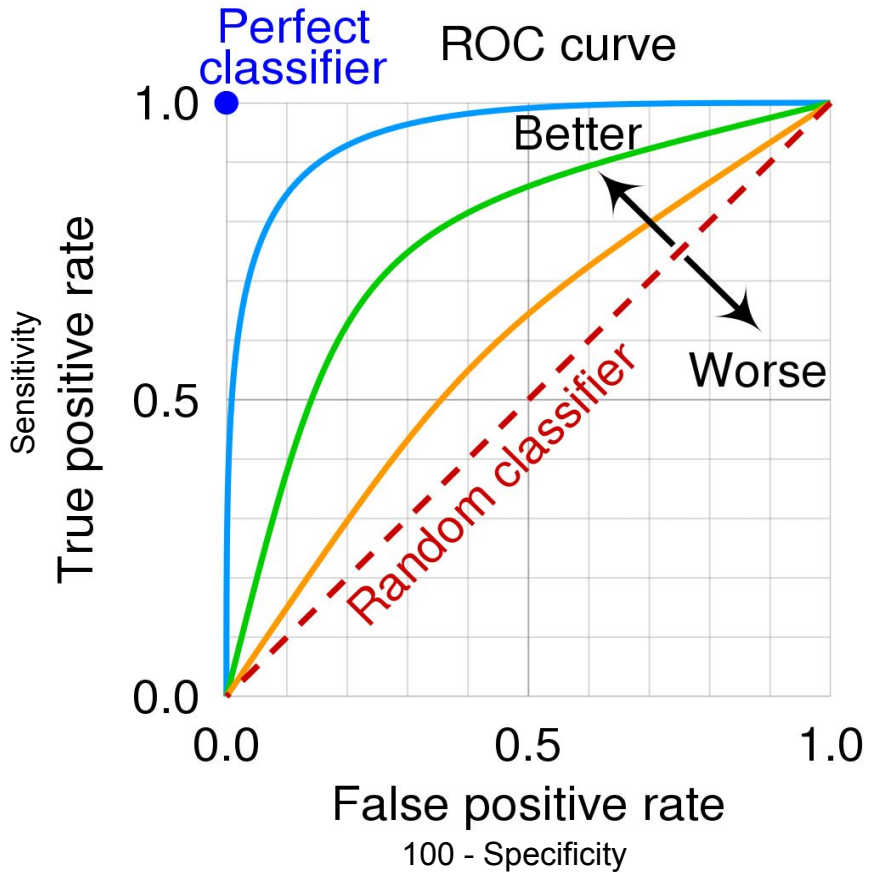
If we change threshold t, how does Sensitivity and Specificity change?

7. ROC and AUC


Perfect classifier
ROC curve
1.0
Better
Worse
0.5
Random classifier
Sensitivity
True positive rate
0.0
0.0
0.5
1.0
False positive rate
100 - Specificity

7. ROC and AUC

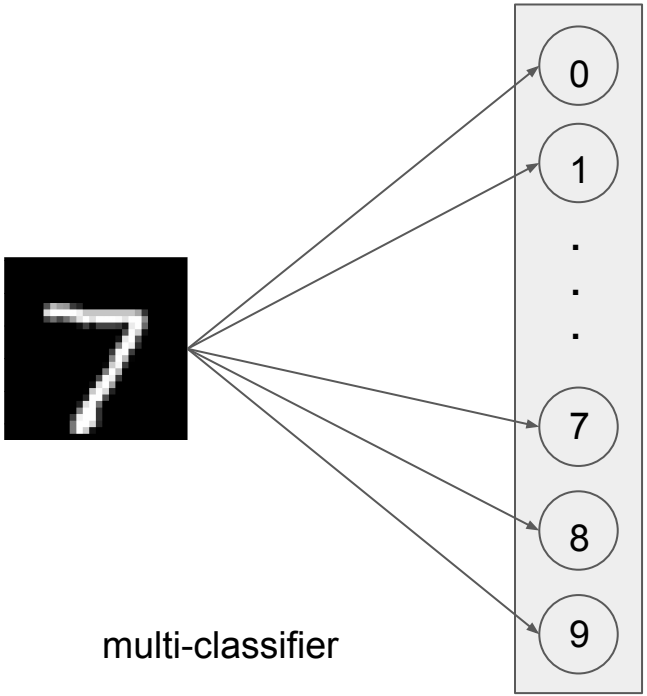
- *AUC = Area under the curve ROC*
- *AUC in [0, 1.0]*
- *Bigger AUC is greater classifier*

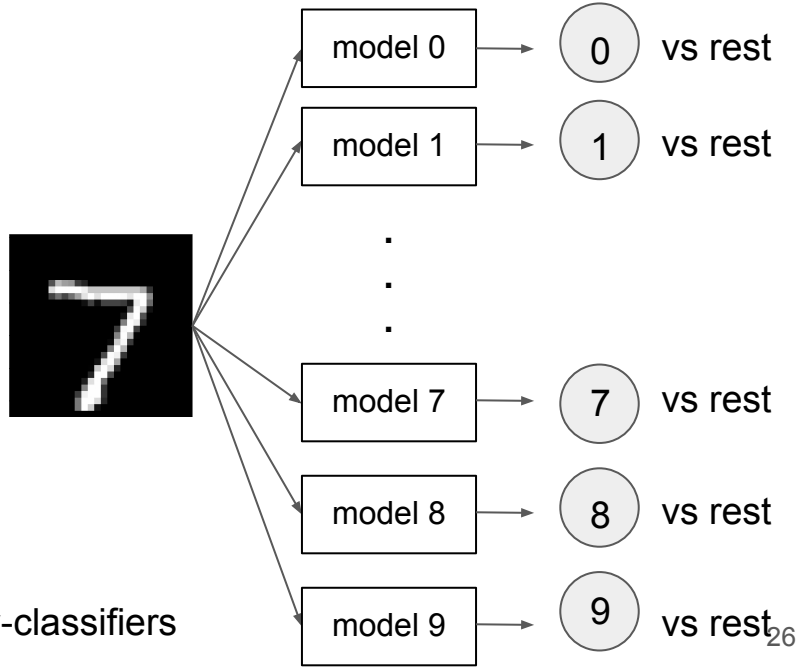


8. Evaluate multi-classification model

DS boosting your data exploration

- Multi-classifier literally is a group of binary classifier
- One-vs-rest method:





8. Evaluate multi-classification model

DS boosting your data exploration

- Evaluate class *i* based on binary-classifier *model i*