Decision Theory

Performance evaluation review

Metrics & Evaluation

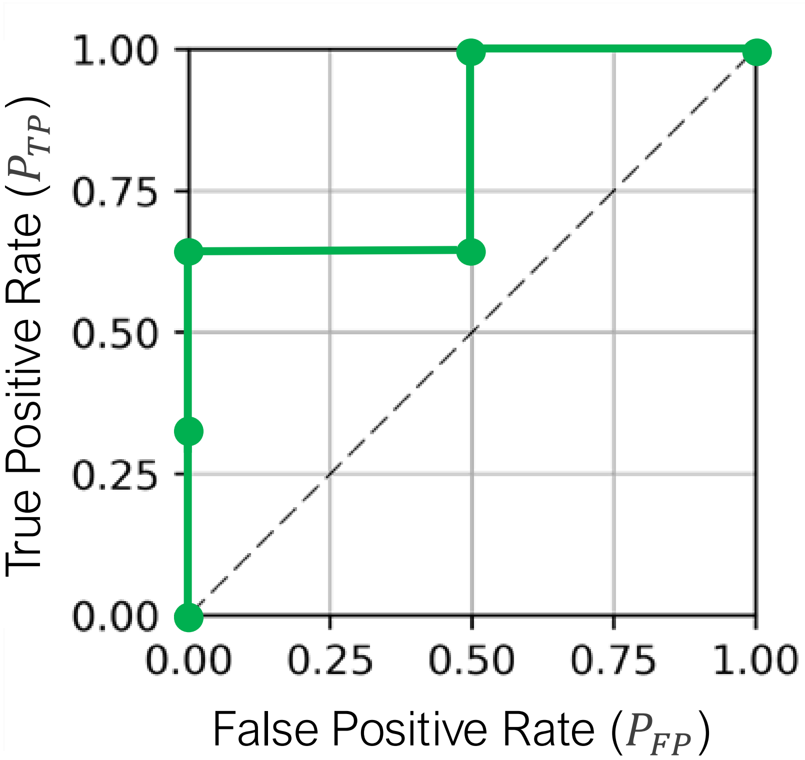
(regression/classification metrics, ROC curves)

Quantify model performance

Experimental Design	Set of decisions to fairly compare models to determine what impacts model performance
Model Comparison	Fairly **compare** model generalization performance
Performance Evaluation	Estimate generalization performance

Performance Evaluation → Application

How do we use this information…



…to make practical predictions?

Time to make a decision…

Exercise inspired by Mausam, University of Washington, CSE573

State of Nature

Action	Poor market performance **Payoff**	Good market performance **Payoff**	
Buy Apple	-1,000	1,700	-10% to +17% return
Buy Google	-2,000	2,100	-20% to +21% return
Buy bonds	500	500	Guaranteed 5% return

How to invest $10,000?

Maximax

Optimism

Select the maximum of the maximum payoff

Action	State of Nature: Poor market performance **Payoff**	State of Nature: Good market performance **Payoff**	Criterion: Maximum payoff for an action
Buy Apple	-1,000	1,700	1,700
Buy Google	-2,000	2,100	2,100
Buy bonds	500	500	500

← **Maximax** (Buy Google)

Maximin

Pessimism

Select the maximum of the minimum payoffs

Action	State of Nature: Poor market performance Payoff	State of Nature: Good market performance Payoff	Criterion: Minimum payoff for an action	
Buy Apple	-1,000	1,700	-1,000	
Buy Google	-2,000	2,100	-2,000	
Buy bonds	500	500	500	← **Maximin**

Minimax

Select the minimum maximum regret

	State of Nature				Criterion
	Poor market performance		Good market performance		Maximum regret for an action
	Payoff	**Regret**	**Payoff**	**Regret**	
Buy Apple	-1,000	1,500	1,700	400	1,500 ← **Minimax**
Buy Google	-2,000	2,500	2,100	0	2,500
Buy bonds	500	0	500	1,600	1,600

Action

Which decision would I regret least?

Regret = Opportunity Loss
Difference between a decision made and an optimal decision

Next: factor in probabilities of different outcomes

Expected Payoff: Equal likelihood

Select the highest average payoff ASSUMING all states are of equal probability

Action	State of Nature: Poor market performance Payoff	State of Nature: Good market performance Payoff	Criterion: Expected reward/ payoff
Buy Apple	-1,000	1,700	350
Buy Google	-2,000	2,100	50
Buy bonds	500	500	500
State Probability:	0.5	0.5	

← **Maximum Expected Reward**

Expected Payoff

Action	State of Nature: Poor market performance Payoff	State of Nature: Good market performance Payoff	Criterion: Expected reward/ payoff
Buy Apple	-1,000	1,700	890
Buy Google	-2,000	2,100	870
Buy bonds	500	500	500
State Probability:	0.3	0.7	

Select the highest average payoff assuming state probabilities from prior knowledge

Maximum Expected Reward

Decision making design pattern

1. Define a measure of risk or reward
2. Select the action that optimizes that metric

Notation

$EV(a_i) = \lambda(a_i|s_0)P(s_0) + \lambda(a_i|s_1)P(s_1)$

$\uparrow$ Expected reward

State of Nature (s)

Action	Poor market performance $s = s_0$	Excellent market performance $s = s_1$	**Expected Reward** $EV(a_i)$
Buy Apple $a = a_0$	$\lambda(a_0\|s_0)$ -1,000	$\lambda(a_0\|s_1)$ 1,700	$(0.3)(-1000) + (0.7)(1700)$ $= \mathbf{890}$
Buy Google $a = a_1$	$\lambda(a_1\|s_0)$ -2,000	$\lambda(a_1\|s_1)$ 2,100	$(0.3)(-2000) + (0.7)(2100)$ $= \mathbf{870}$
Buy bonds $a = a_2$	$\lambda(a_2\|s_0)$ 500	$\lambda(a_2\|s_1)$ 500	$(0.3)(500) + (0.7)(500)$ $= \mathbf{500}$
State Probability:	$P(s_0) =$ 0.3	$P(s_1) =$ 0.7	

Risk = expected loss
(cost)

Loss: $\lambda(a_i|s_j) \triangleq$ Loss incurred by choosing action i and the state of nature being state j

Risk:
Expected loss

$$R(a_i) = \sum_{j=1}^{N_s} \lambda(a_i|s_j)P(s_j)$$

Goal: Select action i for which $R(a_i)$ is minimum

Payoff

State of Nature

Action		Poor market performance	Good market performance
	Buy Apple	-1,000	1,700
	Buy Google	-2,000	2,100
	Buy bonds	500	500

Loss

(here we define loss in terms of opportunity cost)

State of Nature

Action		Poor market performance	Good market performance
	Buy Apple	1,500	400
	Buy Google	2,500	0
	Buy bonds	0	1,600

Investments: loss

$R(a_i) = \lambda(a_i|s_0)P(s_0) + \lambda(a_i|s_1)P(s_1)$

Risk (Expected loss)

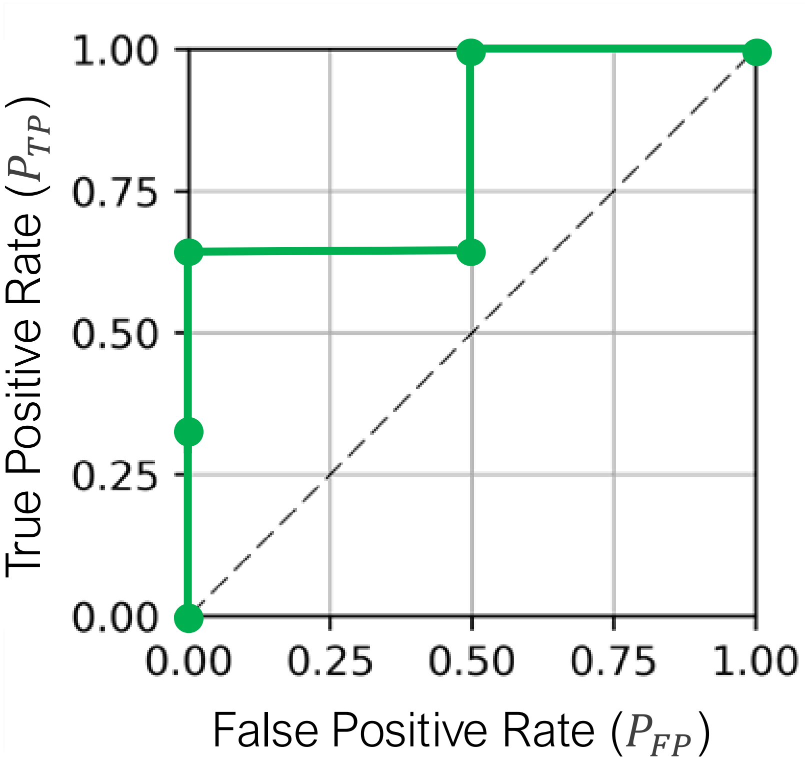
State of Nature (s)

Action	Poor market performance $s = s_0$	Excellent market performance $s = s_1$	**Risk** (Expected Loss) $R(a_i)$
Buy Apple $a = a_0$	$\lambda(a_0\|s_0)$ 1,500	$\lambda(a_0\|s_1)$ 400	$(0.3)(1500) + (0.7)(400) = \mathbf{730}$
Buy Google $a = a_1$	$\lambda(a_1\|s_0)$ 2,500	$\lambda(a_1\|s_1)$ 0	$(0.3)(2500) + (0.7)(0) = \mathbf{750}$
Buy bonds $a = a_2$	$\lambda(a_2\|s_0)$ 0	$\lambda(a_2\|s_1)$ 1,600	$(0.3)(0) + (0.7)(1600) = \mathbf{1220}$

State Probability: $P(s_0) =$ 0.3 $P(s_1) =$ 0.7

We can use risk to choose where to operate along an ROC curve

Performance Evaluation → Application



The "actions" we can evaluate here are which point along the ROC curve to operate?

In other words – what **threshold do I pick** for my classifier decision rule?

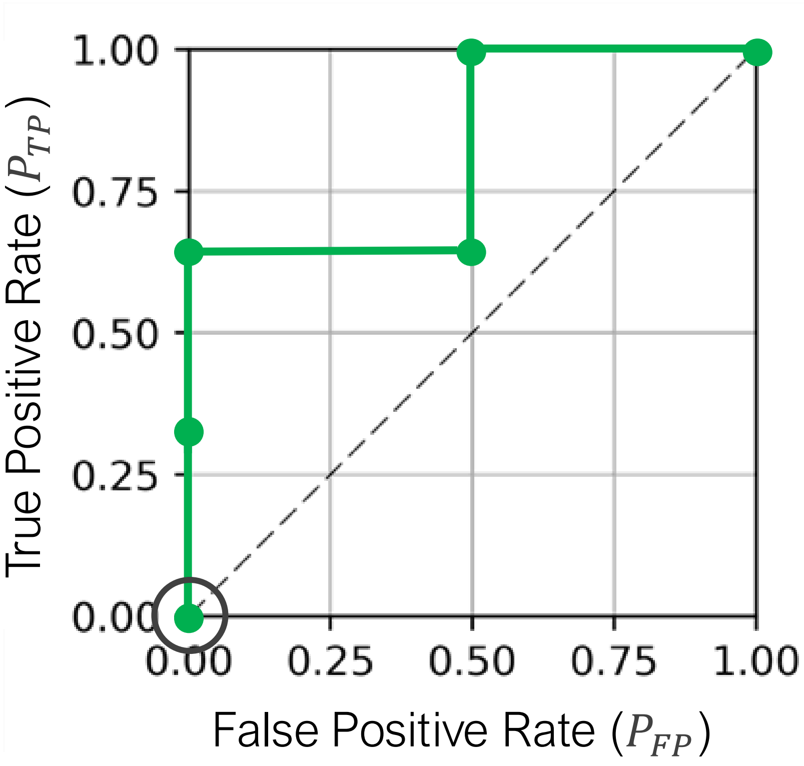
action

Classifier decision rule:

$$\hat{y} = \begin{cases} 1, & \text{confidence score} > \text{thresh} \\ 0, & \text{confidence score} \leq \text{thresh} \end{cases}$$

Note: we define: $\tau \equiv \text{thresh}$

Performance Evaluation → Application



Each point along the ROC represents a different confusion matrix and different rates for TP, FP, TN, FN

Raw values

Predicted Class

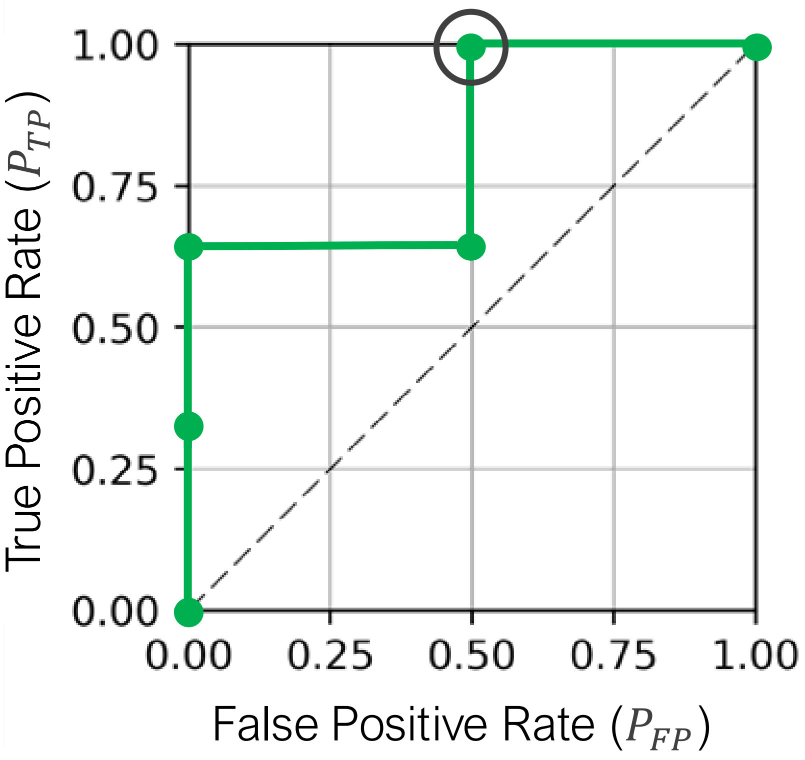
True Class	Class 1	Class 0	
Class 1	TP **0**	FN **50**	[50]
Class 0	FP **0**	TN **50**	[50]

Rates / probabilities

Predicted Class

True Class	Class 1	Class 0	
Class 1	TPR **0**	FNR **1.0**	[50]
Class 0	FPR **0**	TNR **1.0**	[50]

Performance Evaluation → Application



Each point along the ROC represents a different confusion matrix and different rates for TP, FP, TN, FN

Raw values

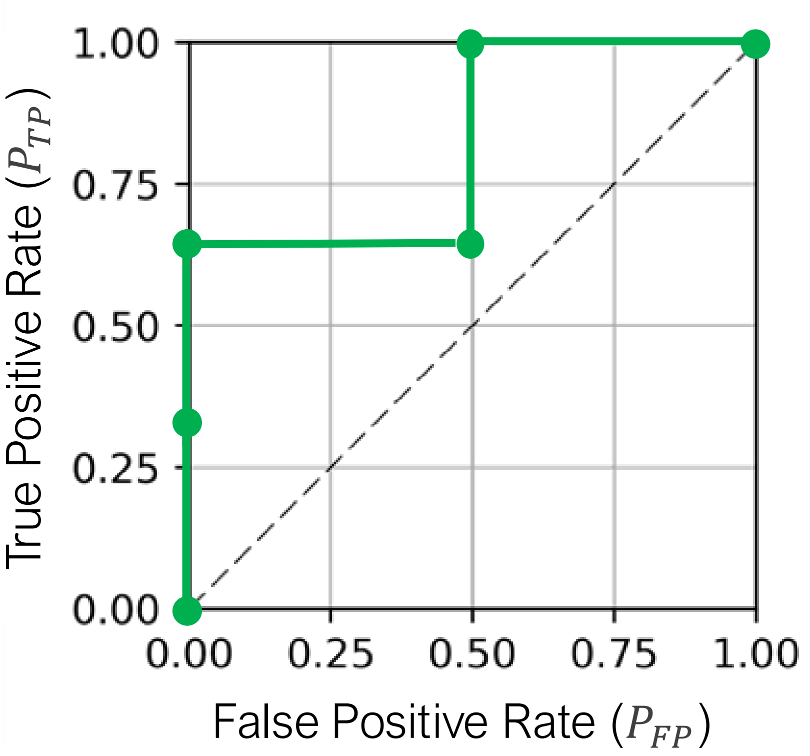
		Predicted Class: Class 1	Predicted Class: Class 0	
True Class	Class 1	TP **50**	FN **0**	[50]
True Class	Class 0	FP **25**	TN **25**	[50]

Rates / probabilities

		Predicted Class: Class 1	Predicted Class: Class 0	
True Class	Class 1	TPR **1.0**	FNR **0**	[50]
True Class	Class 0	FPR **0.5**	TNR **0.5**	[50]

Where to operate along ROC?

		Estimate	
		Class 1	Class 0
Truth	Class 1	**True Positive** $\lambda_{11} = \mathbf{0}$	**False negative** $\lambda_{01} = \mathbf{100}$
	Class 0	**False Positive** $\lambda_{10} = \mathbf{1}$	**True Negative** $\lambda_{00} = \mathbf{0}$



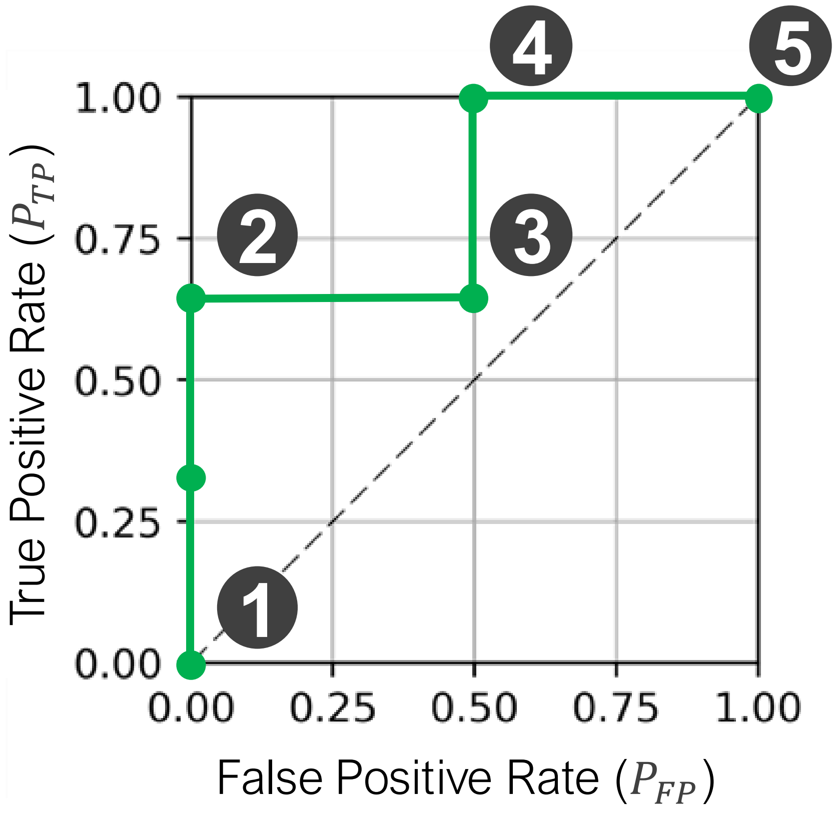
$\lambda_{ij} = \lambda(\hat{y} = i, y = j)$

Loss if prediction is class i when truth is class j

- Assume for our problem a false negative is 100 times worse than a false positive
- Correct predictions incur no penalties

Choose where to operate

Action: select threshold k	Probability of false positive P_{FP}	Probability of false negative $P_{FN} = (1 - P_{TP})$	Risk $R(\tau_k)$
1			
2			
3			
4			
5			



$$R(\tau_k) = \sum_{i,j} \lambda(\hat{y} = i, y = j) P(\hat{Y} = i | Y = j, \tau_k)$$

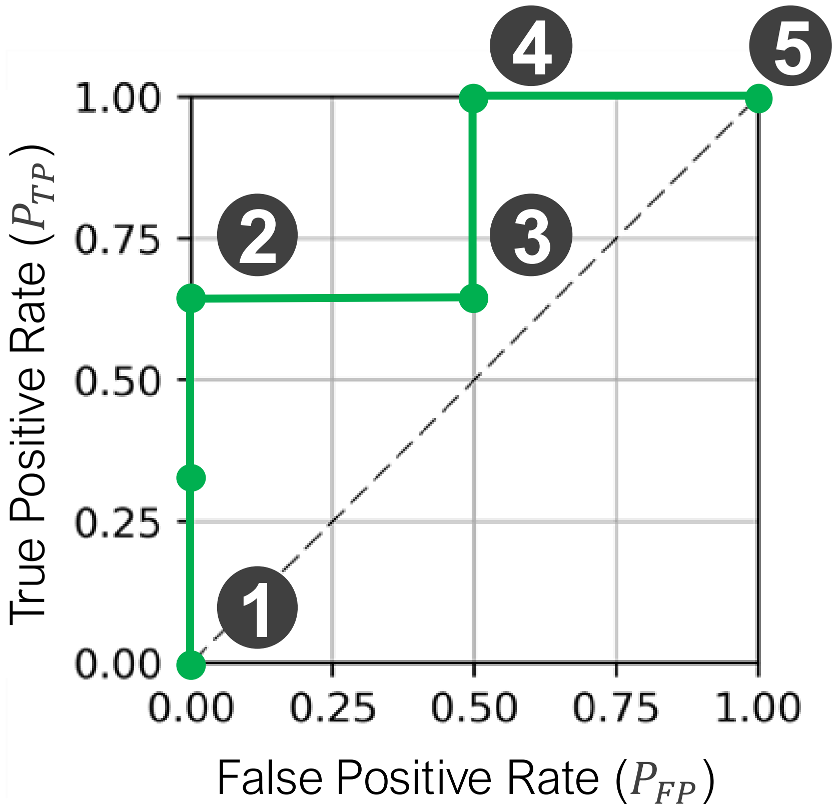
$$R(\tau_k) = \lambda_{10} P_{FP}(\tau_k) + \lambda_{01} P_{FN}(\tau_k)$$

Estimate

Truth	Class 1	Class 0
Class 1	**True Positive** $\lambda_{11} = 0$	**False negative** $\lambda_{01} = 100$
Class 0	**False Positive** $\lambda_{10} = 1$	**True Negative** $\lambda_{00} = 0$

Choose where to operate

Action: select threshold k	Probability of false positive P_{FP}	Probability of false negative $P_{FN} = (1 - P_{TP})$	Risk $R(\tau_k)$
1	0	1	100
2			
3			
4			
5			



$$R(\tau_k) = \sum_{i,j} \lambda(\hat{y} = i, y = j) P(\hat{Y} = i | Y = j, \tau_k)$$

$$R(\tau_k) = \lambda_{10} P_{FP}(\tau_k) + \lambda_{01} P_{FN}(\tau_k)$$

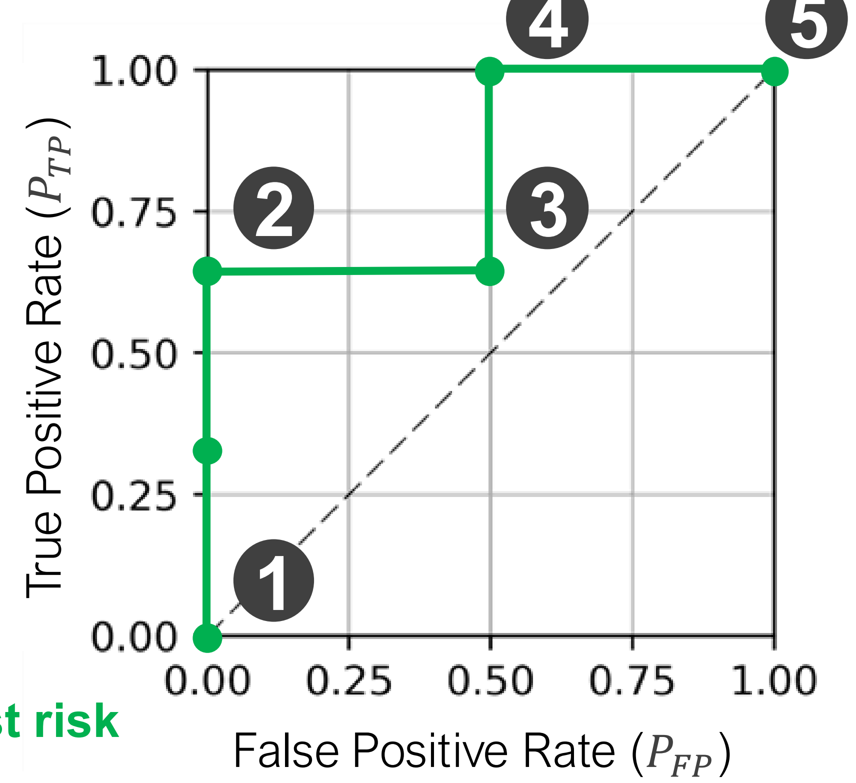
Estimate

Truth	Class 1	Class 0
Class 1	**True Positive** $\lambda_{11} = \mathbf{0}$	**False negative** $\lambda_{01} = \mathbf{100}$
Class 0	**False Positive** $\lambda_{10} = \mathbf{1}$	**True Negative** $\lambda_{00} = \mathbf{0}$

Choose where to operate

Action: select threshold k	Probability of false positive P_{FP}	Probability of false negative $P_{FN} = (1 - P_{TP})$	Risk $R(\tau_k)$
1	0	1	100
2	0	0.33	33
3	0.5	0.33	33.5
4	0.5	0	0.5
5	1	0	1

Least risk



$$R(\tau_k) = \sum_{i,j} \lambda(\hat{y} = i, y = j) P(\hat{Y} = i | Y = j, \tau_k)$$

$$R(\tau_k) = \lambda_{10} P_{FP}(\tau_k) + \lambda_{01} P_{FN}(\tau_k)$$

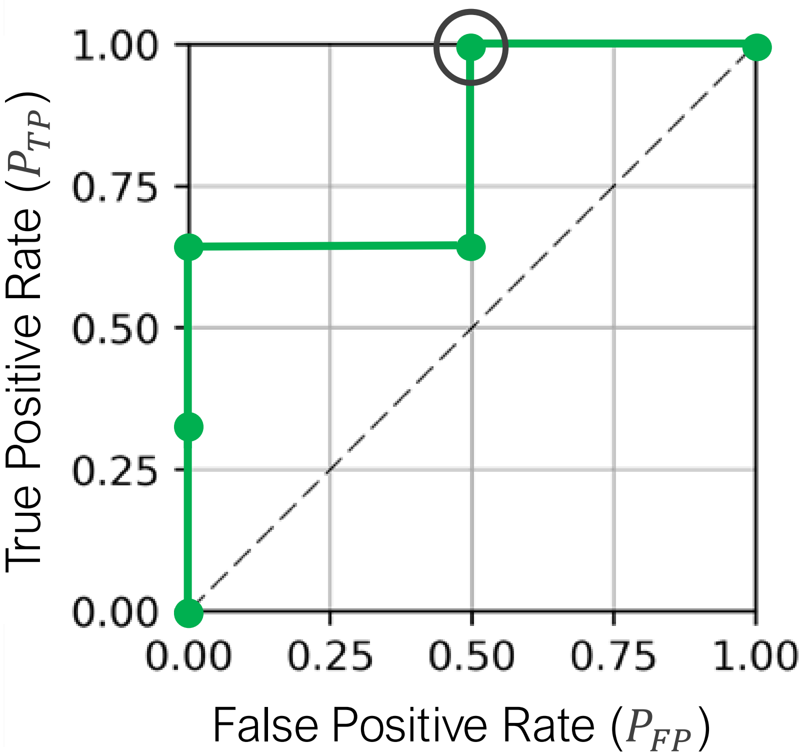
Estimate

Truth	Class 1	Class 0
Class 1	True Positive $\lambda_{11} = 0$	False negative $\lambda_{01} = 100$
Class 0	False Positive $\lambda_{10} = 1$	True Negative $\lambda_{00} = 0$

This makes a critical assumption about how the data are distributed by class…

Note: in the previous examples, if we wanted risk to equal expected loss, we would need to have assigned a probability of 0.5 to each class

Performance Evaluation → Application



Each point along the ROC represents a different confusion matrix and different rates for TP, FP, TN, FN

Raw values

Predicted Class

True Class	Class 1	Class 0	
Class 1	TP **50**	FN **0**	[50]
Class 0	FP **25**	TN **25**	[50]

Rates / probabilities

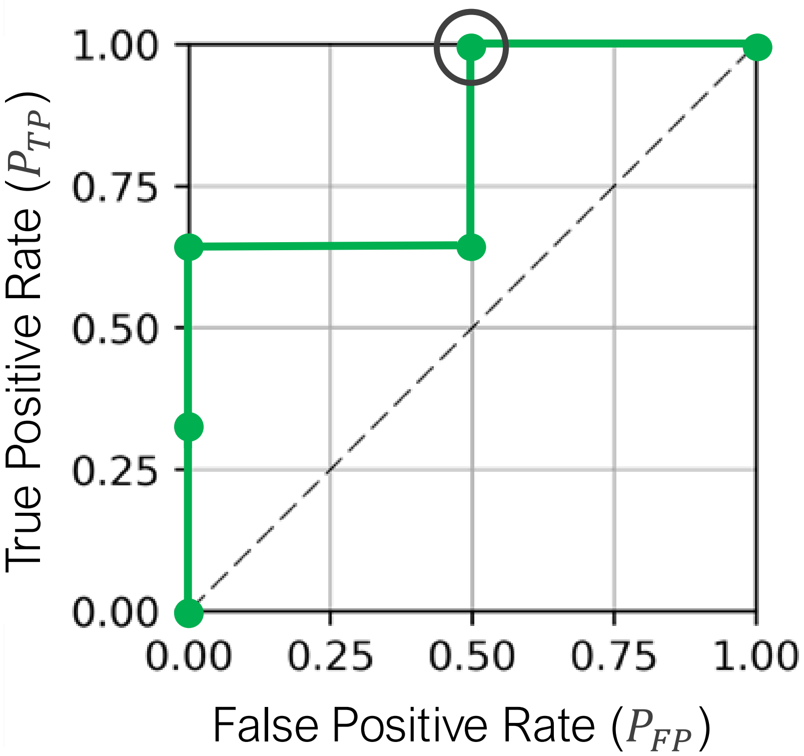
Predicted Class

True Class	Class 1	Class 0	
Class 1	TPR **1.0**	FNR **0**	[50]
Class 0	FPR **0.5**	TNR **0.5**	[50]

Precision = 50 / 75 = 0.67

Classes are equiprobable
$P(Y = 1) = P(Y = 0) \cong 0.5$

Performance Evaluation → Application



Each point along the ROC represents a different confusion matrix and different rates for TP, FP, TN, FN

Raw values

Predicted Class

True Class		Class 1	Class 0	
	Class 1	TP **5**	FN **0**	[5]
	Class 0	FP **25**	TN **25**	[50]

Rates / probabilities

Predicted Class

True Class		Class 1	Class 0	
	Class 1	TPR **1.0**	FNR **0**	[5]
	Class 0	FPR **0.5**	TNR **0.5**	[50]

Precision = 5 / 30 = 0.17

$$P(Y=1) \cong \frac{5}{55} = 0.09$$

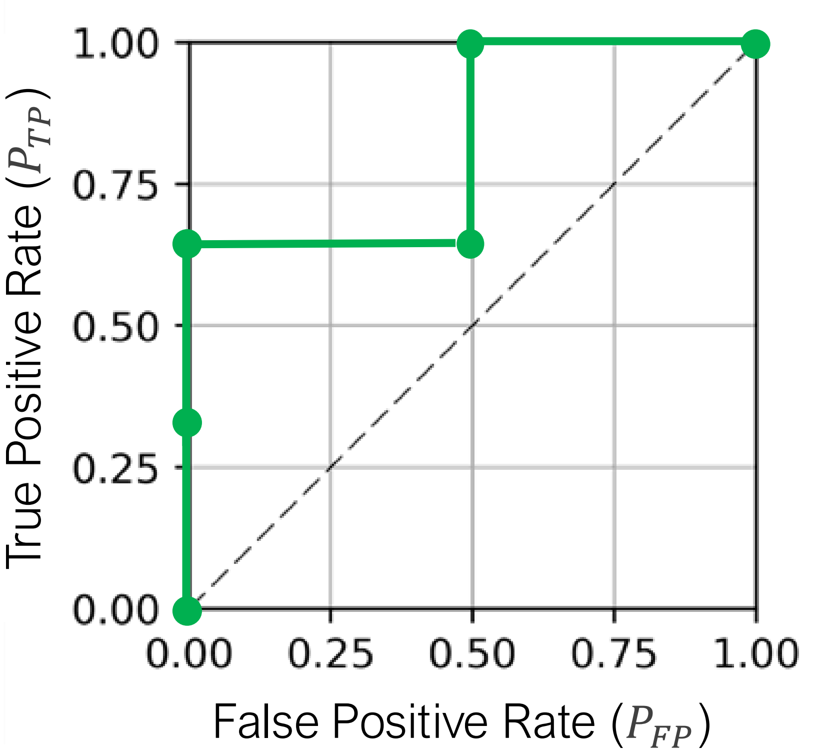
$$P(Y=0) \cong \frac{50}{55} = 0.89$$

ROC curves are not sensitive to class imbalances – TPR and FPR are class-conditional

This means we need to factor in the prevalence of each class

Where to operate along ROC?

		Estimate	
		Class 1	Class 0
Truth	Class 1 $\pi_1 = P(Y=1)$	**True Positive** $\lambda_{11} = 0$	**False negative** $\lambda_{01} = 100$
	Class 0 $\pi_0 = P(Y=0)$	**False Positive** $\lambda_{10} = 1$	**True Negative** $\lambda_{00} = 0$



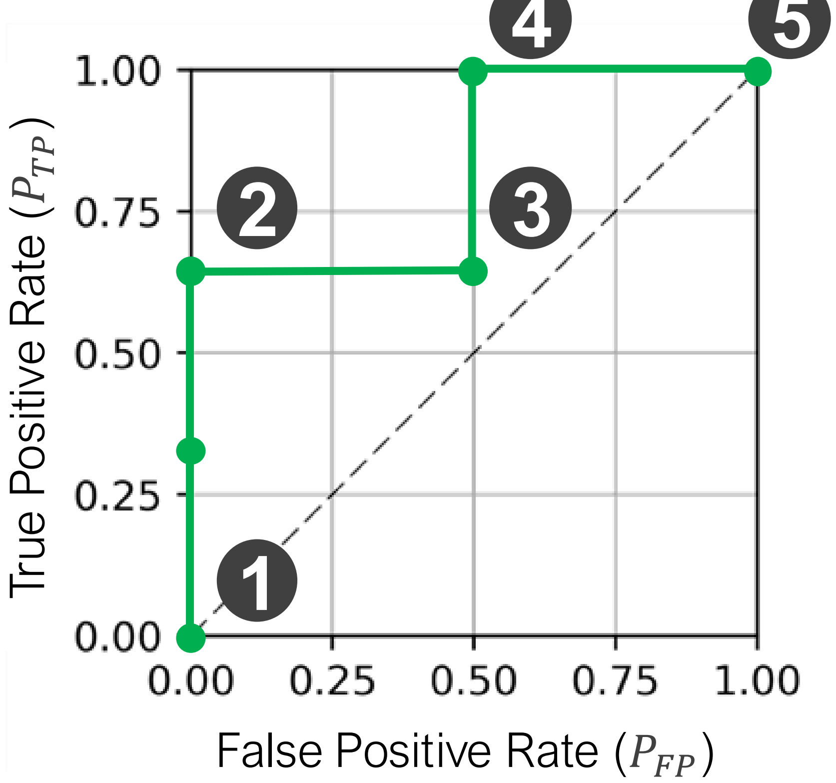
$$\lambda_{ij} = \lambda(\hat{y} = i, y = j)$$

Loss if prediction is class i when truth is class j

Class prevalence also needs to be factored in!

Choose where to operate

Action: select threshold k	Probability of false positive P_{FP}	Probability of false negative $P_{FN} = (1 - P_{TP})$	Risk $R(\tau_k)$
1			
2			
3			
4			
5			



$P(Y = 0) \cong 0.99$
$P(Y = 1) \cong 0.01$

		Estimate	
		Class 1	Class 0
Truth	Class 1	True Positive $\lambda_{11} = 0$	False negative $\lambda_{01} = 100$
	Class 0	False Positive $\lambda_{10} = 1$	True Negative $\lambda_{00} = 0$

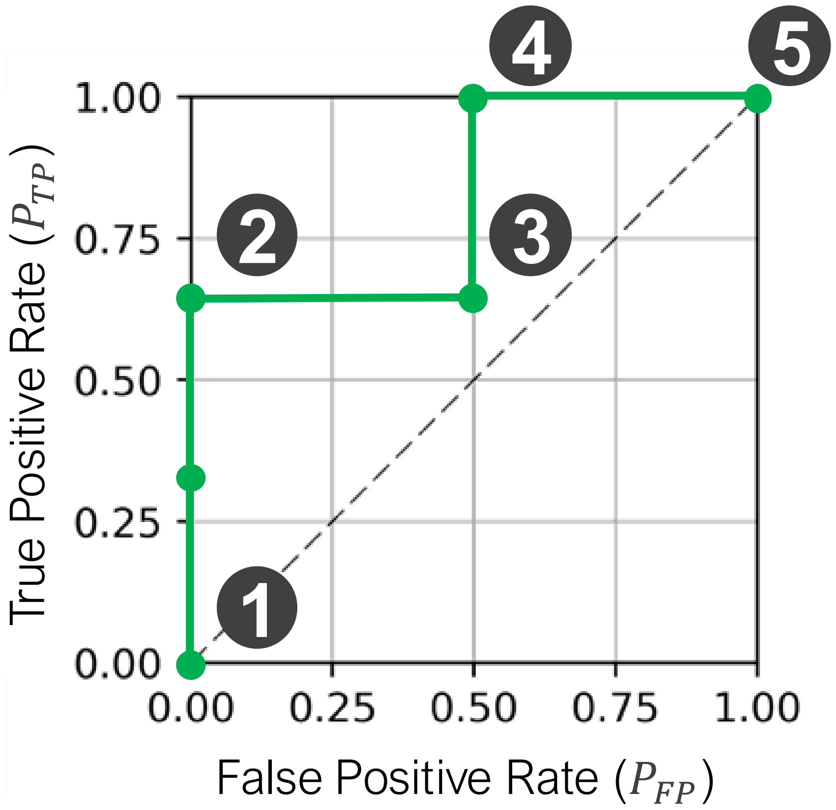
$$R(\tau_k) = \sum_{i,j} \lambda(\hat{y} = i, y = j) P(\hat{Y} = i | Y = j, \tau_k) P(Y = j)$$

$$\pi_i = P(Y = i)$$

$$R(\tau_k) = \lambda_{10} P_{FP}(\tau_k) \pi_0 + \lambda_{01} P_{FN}(\tau_k) \pi_1$$

Choose where to operate

Action: select threshold k	Probability of false positive P_{FP}	Probability of false negative $P_{FN} = (1 - P_{TP})$	Risk $R(\tau_k)$
1	0	1	
2	0	0.33	
3	0.5	0.33	
4	0.5	0	
5	1	0	



$P(Y = 0) \cong 0.99$
$P(Y = 1) \cong 0.01$

Estimate

Truth	Class 1	Class 0
Class 1	True Positive $\lambda_{11} = 0$	False negative $\lambda_{01} = 100$
Class 0	False Positive $\lambda_{10} = 1$	True Negative $\lambda_{00} = 0$

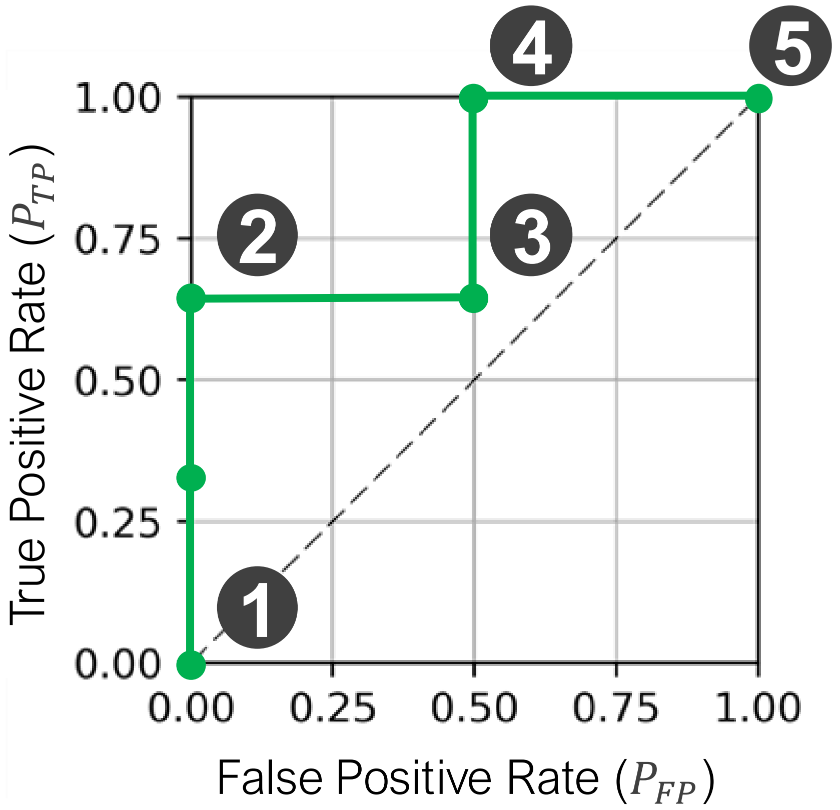
$$R(\tau_k) = \sum_{i,j} \lambda(\hat{y} = i, y = j) P(\hat{Y} = i | Y = j, \tau_k) P(Y = j)$$

$$\pi_i = P(Y = i)$$

$$R(\tau_k) = \lambda_{10} P_{FP}(\tau_k) \pi_0 + \lambda_{01} P_{FN}(\tau_k) \pi_1$$

Choose where to operate

Action: select threshold k	Probability of false positive P_{FP}	Probability of false negative $P_{FN} = (1 - P_{TP})$	Risk $R(\tau_k)$
1	0	1	1
2	0	0.33	
3	0.5	0.33	
4	0.5	0	
5	1	0	



$P(Y = 0) \cong 0.99$
$P(Y = 1) \cong 0.01$

$$R(\tau_k) = \sum_{i,j} \lambda(\hat{y} = i, y = j) P\left(\hat{Y} = i \middle| Y = j, \tau_k\right) P(Y = j)$$

$$\pi_i = P(Y = i)$$

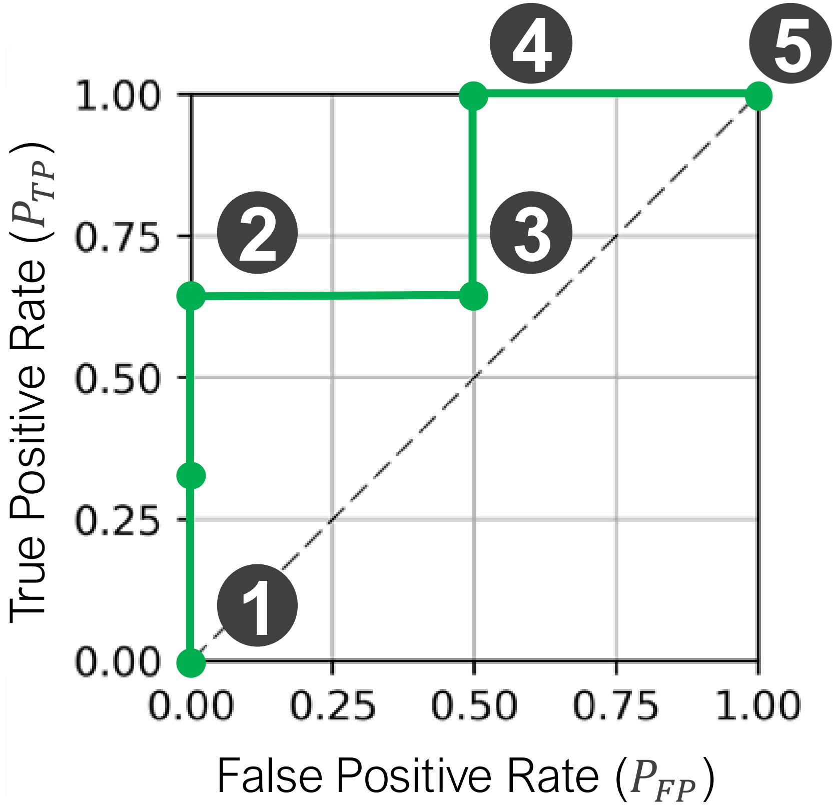
$$R(\tau_k) = \lambda_{10} P_{FP}(\tau_k)\pi_0 + \lambda_{01} P_{FN}(\tau_k)\pi_1$$

Estimate

Truth	Class 1	Class 0
Class 1	**True Positive** $\lambda_{11} = 0$	**False negative** $\lambda_{01} = 100$
Class 0	**False Positive** $\lambda_{10} = 1$	**True Negative** $\lambda_{00} = 0$

Choose where to operate

Action: select threshold k	Probability of false positive P_{FP}	Probability of false negative $P_{FN} = (1 - P_{TP})$	Risk $R(\tau_k)$
1	0	1	1
2	0	0.33	0.33
3	0.5	0.33	
4	0.5	0	
5	1	0	



$P(Y=0) \cong 0.99$
$P(Y=1) \cong 0.01$

$$R(\tau_k) = \sum_{i,j} \lambda(\hat{y}=i, y=j) P(\hat{Y}=i \mid Y=j, \tau_k) P(Y=j)$$

$$\pi_i = P(Y=i)$$

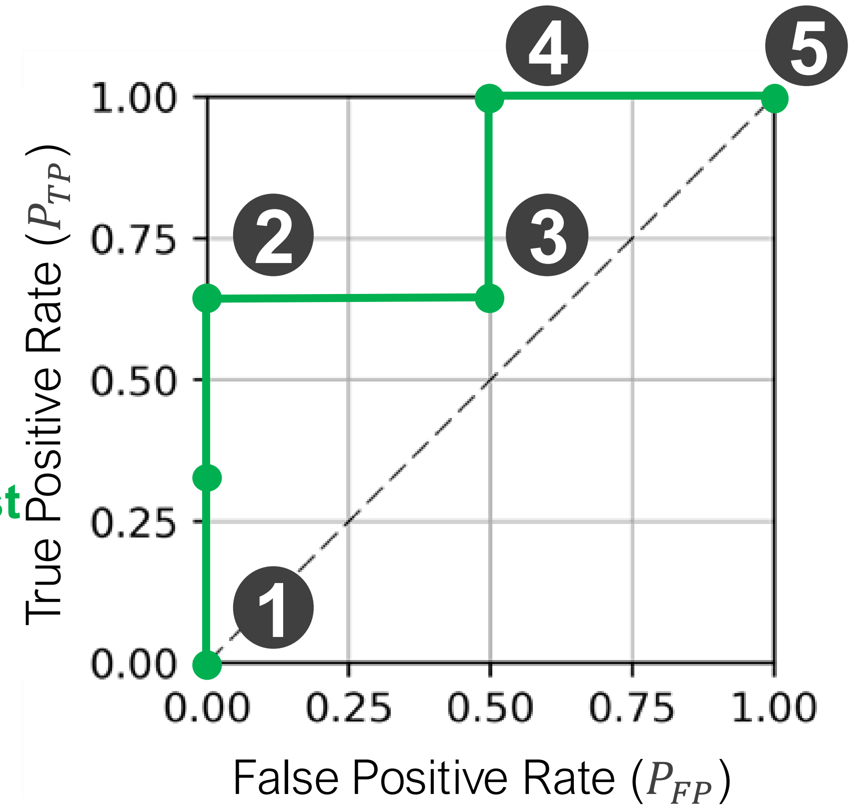
$$R(\tau_k) = \lambda_{10} P_{FP}(\tau_k)\pi_0 + \lambda_{01} P_{FN}(\tau_k)\pi_1$$

Truth	Estimate: Class 1	Estimate: Class 0
Class 1	True Positive $\lambda_{11} = 0$	False negative $\lambda_{01} = 100$
Class 0	False Positive $\lambda_{10} = 1$	True Negative $\lambda_{00} = 0$

Choose where to operate

Action: select threshold k	Probability of false positive P_{FP}	Probability of false negative $P_{FN} = (1 - P_{TP})$	Risk $R(\tau_k)$
1	0	1	1
2	0	0.33	0.33
3	0.5	0.33	0.83
4	0.5	0	0.50
5	1	0	0.99

Least risk



$P(Y = 0) \cong 0.99$
$P(Y = 1) \cong 0.01$

		Estimate Class 1	Class 0
Truth	Class 1	True Positive $\lambda_{11} = 0$	False negative $\lambda_{01} = 100$
	Class 0	False Positive $\lambda_{10} = 1$	True Negative $\lambda_{00} = 0$

$$R(\tau_k) = \sum_{i,j} \lambda(\hat{y} = i, y = j) P\left(\hat{Y} = i \middle| Y = j, \tau_k\right) P(Y = j)$$

$$\pi_i = P(Y = i)$$

$$R(\tau_k) = \lambda_{10} P_{FP}(\tau_k) \pi_0 + \lambda_{01} P_{FN}(\tau_k) \pi_1$$

Takeaways

Models make predictions, but these need to be contextualized to transform them into a decision

To make a decision, we minimize expected loss (risk):

1. Define a measure of risk
2. Select the action that optimizes that metric (for binary classification, this is often the decision rule threshold)

Decision theory systematically incorporates the relative importance of different error types and the prevalence of the types of classes

Requires application domain knowledge to reflect real-world importance