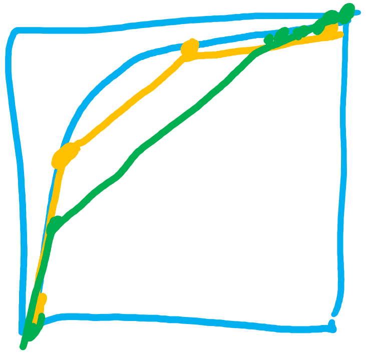
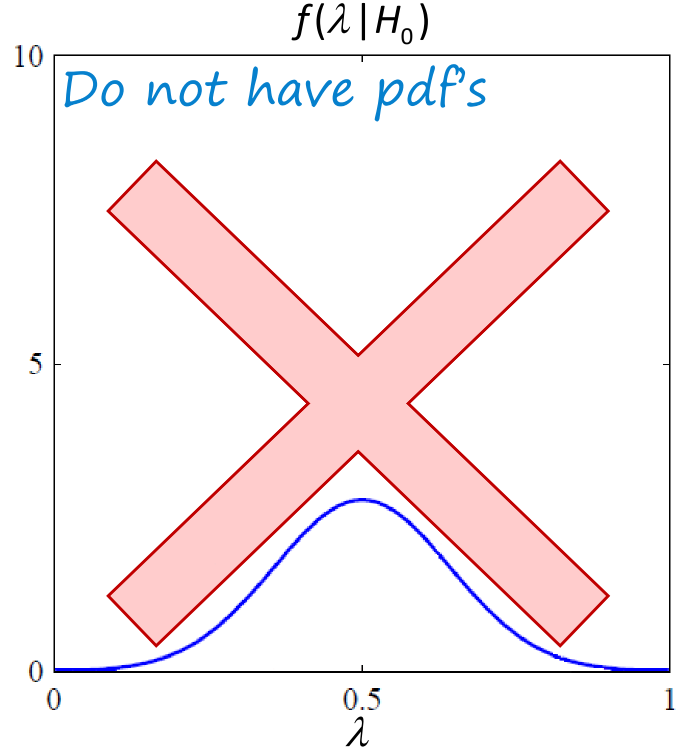
Generating ROCs in Practice

$$P_{CR}(\beta) = \frac{\#H_0 \ \lambda\text{'s} < \beta}{\#H_0 \ \lambda\text{'s}} \qquad P_M(\beta) = \frac{\#H_1 \ \lambda\text{'s} < \beta}{\#H_1 \ \lambda\text{'s}}$$

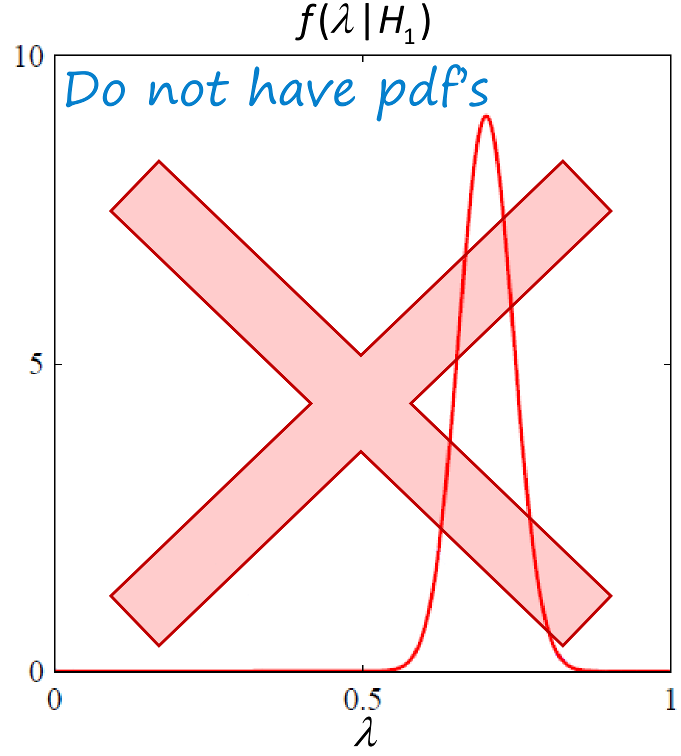
$$P_{FA}(\beta) = \frac{\#H_0 \ \lambda\text{'s} \geq \beta}{\#H_0 \ \lambda\text{'s}} \qquad P_D(\beta) = \frac{\#H_1 \ \lambda\text{'s} \geq \beta}{\#H_1 \ \lambda\text{'s}}$$





Have a list of λ's

λ	truth
0.56	0
0.35	0
0	0
0.21	0
0.11	0



Have a list of λ's

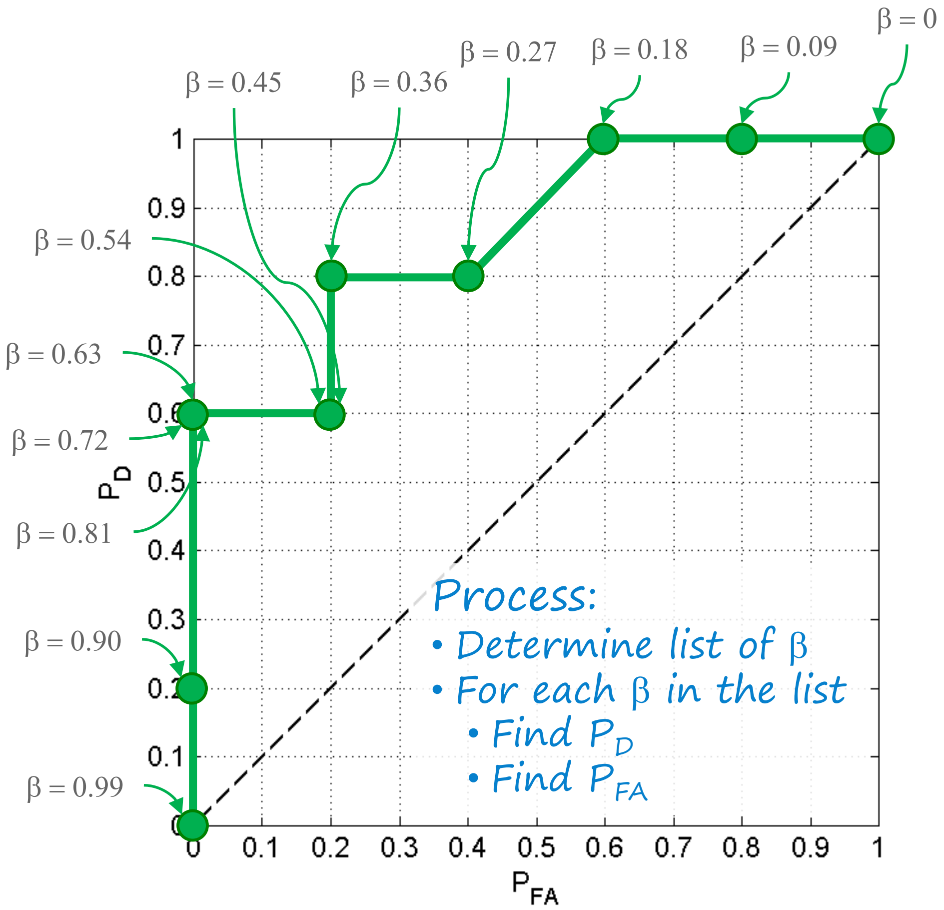
λ	truth
0.18	1
0.92	1
0.42	1
0.88	1
0.82	1

How Should We Select Decision Thresholds?

[linearly span from min(λ) to max(λ)]

Need decision thresholds that span the range of decision statistics

λ	truth	β = 0	β = 0.09	β = 018	β = 0.27	β = 0.36	β = 0.45	β = 0.54	β = 0.63	β = 0.72	β = 0.81	β = 0.90	β = 0.99
0.18	1	✓	✓	✓	--	--	--	--	--	--	--	--	--
0.56	0	✓	✓	✓	✓	✓	✓	✓	--	--	--	--	--
0.35	0	✓	✓	✓	✓	--	--	--	--	--	--	--	--
H_1 0.92	1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	--
H_0 0	0	✓	--	--	--	--	--	--	--	--	--	--	--
0.42	1	✓	✓	✓	✓	✓	--	--	--	--	--	--	--
0.88	1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	--	--
0.82	1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	--	--
0.21	0	✓	✓	✓	--	--	--	--	--	--	--	--	--
0.11	0	✓	✓	--	--	--	--	--	--	--	--	--	--
	P_D	5/5	5/5	5/5	4/5	4/5	3/5	3/5	3/5	3/5	3/5	1/5	0/5
	P_{FA}	5/5	4/5	3/5	2/5	1/5	1/5	1/5	0/5	0/5	0/5	0/5	0/5



Process:
- *Determine list of β*
- *For each β in the list*
  - *Find P_D*
  - *Find P_{FA}*

ECE 580
Spring 2022
Stacy Tantum

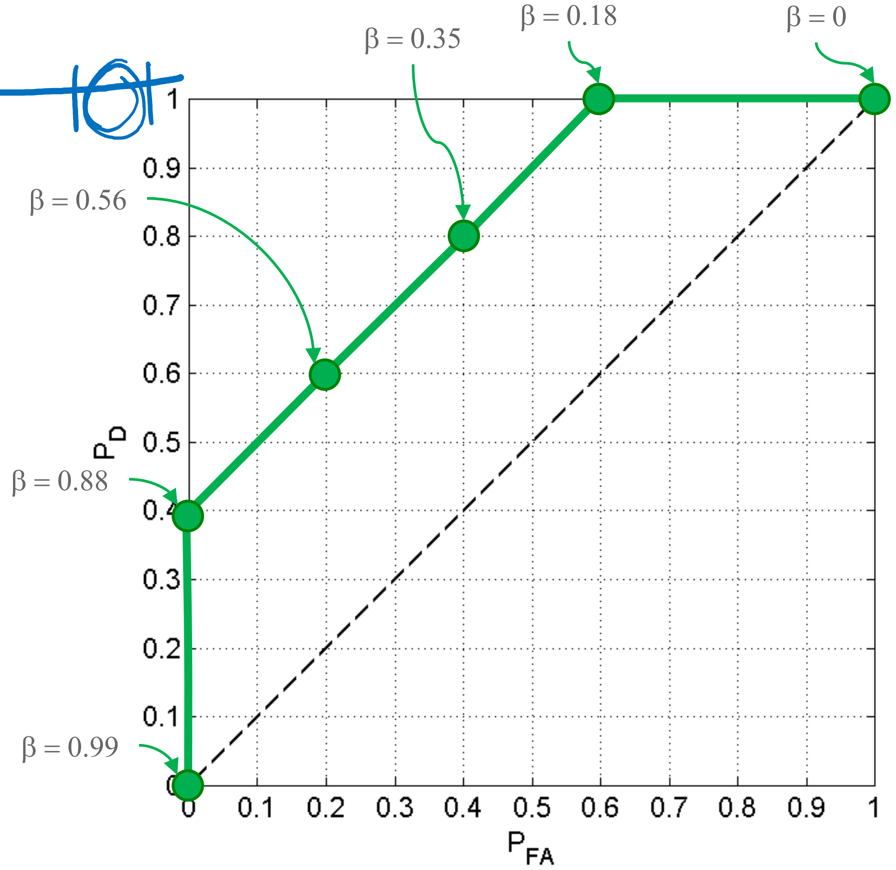
T06.2: Binary Classifier Performance Evaluation

How Should We Select Decision Thresholds?

[sort λ's, choose every n^{th} λ as a β (n=1 for finest resolution]

Need decision thresholds that span the range of decision statistics AND appropriately sample the decision statistics

λ	truth	β = 0	β = 018	β = 0.35	β = 0.56	β = 0.88	β = 0.99
0	0	✓	--	--	--	--	--
0.11	0	✓	--	--	--	--	--
0.18	1	✓	✓	--	--	--	--
0.21	0	✓	✓	--	--	--	--
0.35	0	✓	✓	✓	--	--	--
0.42	1	✓	✓	✓	--	--	--
0.56	0	✓	✓	✓	✓	--	--
0.82	1	✓	✓	✓	✓	--	--
0.88	1	✓	✓	✓	✓	✓	--
0.92	1	✓	✓	✓	✓	✓	--
	P_D	5/5	5/5	4/5	3/5	2/5	0/5
	P_{FA}	5/5	3/5	2/5	1/5	0/5	0/5



ECE 580
Spring 2022
Stacy Tantum

T06.2: Binary Classifier Performance Evaluation

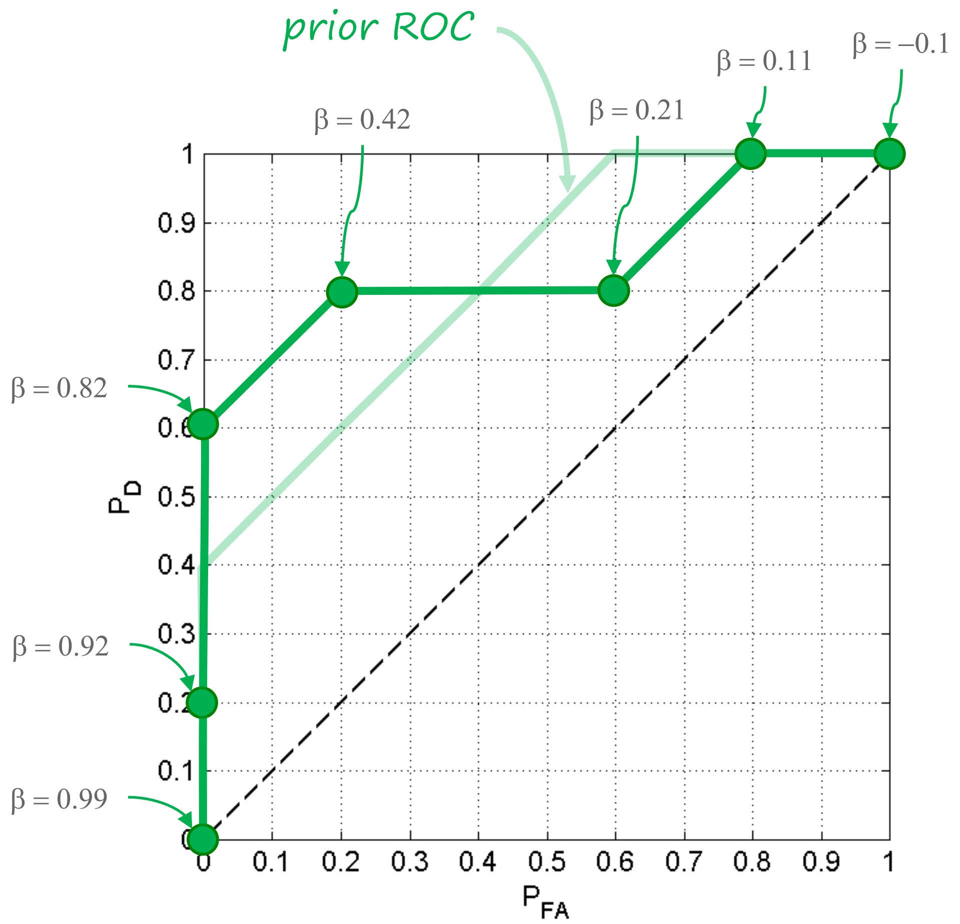
How Should We Select Decision Thresholds?

[sort λ's, choose every n^{th} λ as a β (n=1 for finest resolution)]

Need decision thresholds that span the range of decision statistics AND appropriately sample the decision statistics

λ	truth	β = –0.1	β = 0.11	β = 0.21	β = 0.42	β = 0.82	β = 0.92	β = 0.99
-0.1								
0	0	✓	--	--	--	--	--	--
0.11	0	✓	✓	--	--	--	--	--
0.18	1	✓	✓	--	--	--	--	--
0.21	0	✓	✓	✓	--	--	--	--
0.35	0	✓	✓	✓	--	--	--	--
0.42	1	✓	✓	✓	✓	--	--	--
0.56	0	✓	✓	✓	✓	--	--	--
0.82	1	✓	✓	✓	✓	✓	--	--
0.88	1	✓	✓	✓	✓	✓	--	--
0.92	1	✓	✓	✓	✓	✓	✓	--
	P_D	5/5	5/5	4/5	4/5	3/5	1/5	0/5
	P_{FA}	5/5	4/5	3/5	1/5	0/5	0/5	0/5

add β=-0.1 to complete the ROC



Duke ELECTRICAL & COMPUTER ENGINEERING

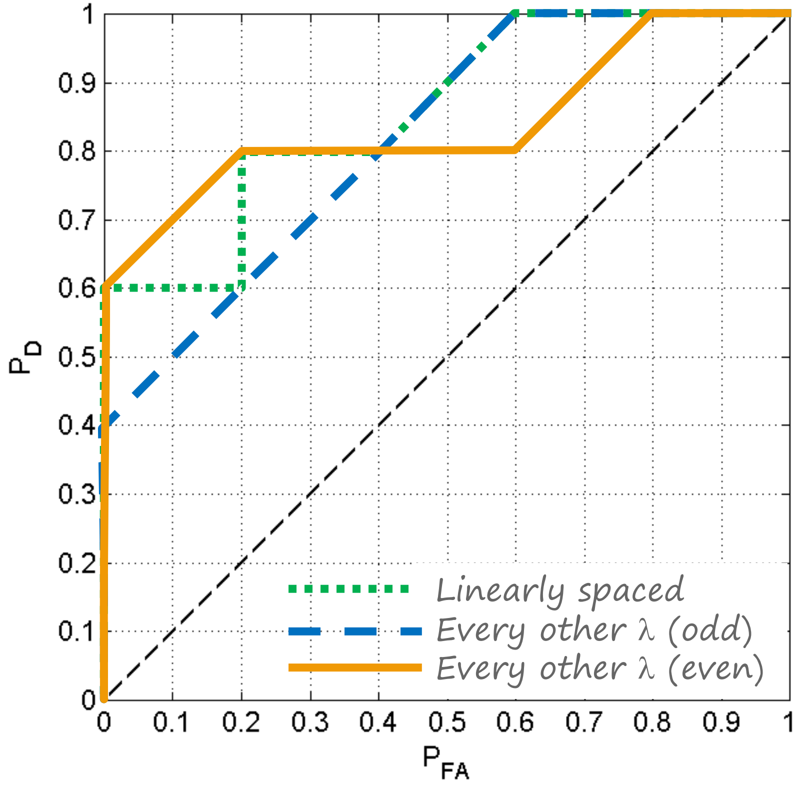
ECE 580 Spring 2022 Stacy Tantum

T06.2: Binary Classifier Performance Evaluation

How Should We Select Decision Thresholds?

The ROC may vary a lot depending on how thresholds are selected

λ	truth
0	0
0.11	0
0.18	1
0.21	0
0.35	0
0.42	1
0.56	0
0.82	1
0.88	1
0.92	1



Limited data (small number of observations) amplifies differences between ROCs

ROCs become very similar with large number of data points (large number of observations)

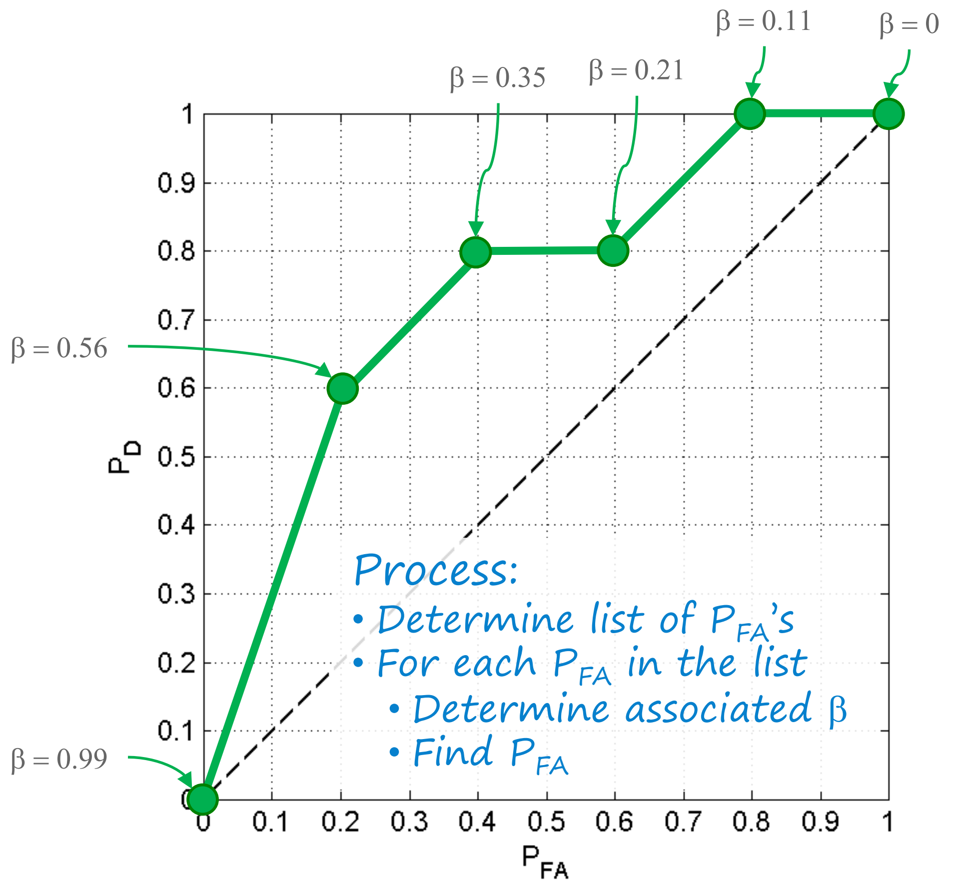
How Should We Select Decision Thresholds?

[sort λ's, choose every n^{th} H_0 λ as a β (n=1 for finest resolution)]

Need decision thresholds that span the range of decision statistics, appropriately sample the decision statistics, AND appropriately sample the ROC

λ	truth	β = 0	β = 0.11	β = 0.21	β = 0.35	β = 0.56	β = 0.99
0	0	✓	--	--	--	--	--
0.11	0	✓	✓	--	--	--	--
0.18	1	✓	✓	--	--	--	--
0.21	0	✓	✓	✓	--	--	--
0.35	0	✓	✓	✓	✓	--	--
0.42	1	✓	✓	✓	✓	--	--
0.56	0	✓	✓	✓	✓	✓	--
0.82	1	✓	✓	✓	✓	✓	--
0.88	1	✓	✓	✓	✓	✓	--
0.92	1	✓	✓	✓	✓	✓	--
0.99	P_D	5/5	5/5	4/5	4/5	3/5	0/5
	P_{FA}	5/5	4/5	3/5	2/5	1/5	0/5

add β=0.99 to complete the ROC



Process:
- Determine list of P_{FA}'s
- For each P_{FA} in the list
  - Determine associated β
  - Find P_{FA}

How Should We Select Decision Thresholds?

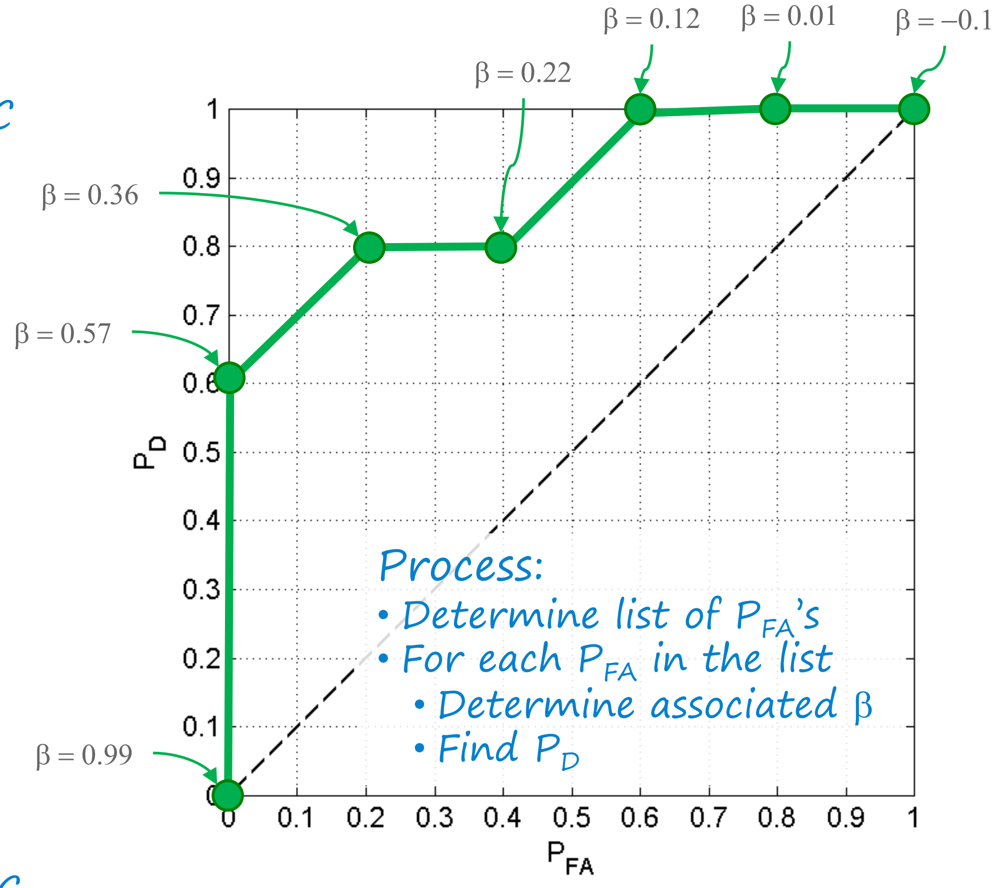
[sort λ's, choose every n^{th} (H_0 λ + eps) as a β (n=1 for finest resolution)]

Need decision thresholds that span the range of decision statistics, appropriately sample the decision statistics, AND appropriately sample the ROC

λ	truth	β = −0.1	β = 0.01	β = 0.12	β = 0.22	β = 0.36	β = 0.57	β = 0.99
-0.1								
0	0	✓	--	--	--	--	--	--
0.11	0	✓	✓	--	--	--	--	--
0.18	1	✓	✓	✓	--	--	--	--
0.21	0	✓	✓	✓	--	--	--	--
0.35	0	✓	✓	✓	✓	--	--	--
0.42	1	✓	✓	✓	✓	✓	--	--
0.56	0	✓	✓	✓	✓	✓	--	--
0.82	1	✓	✓	✓	✓	✓	✓	--
0.88	1	✓	✓	✓	✓	✓	✓	--
0.92	1	✓	✓	✓	✓	✓	✓	--
0.99	P_D	5/5	5/5	5/5	4/5	4/5	3/5	0/5
	P_{FA}	5/5	4/5	3/5	2/5	1/5	0/5	0/5

add β=-0.1 to complete the ROC

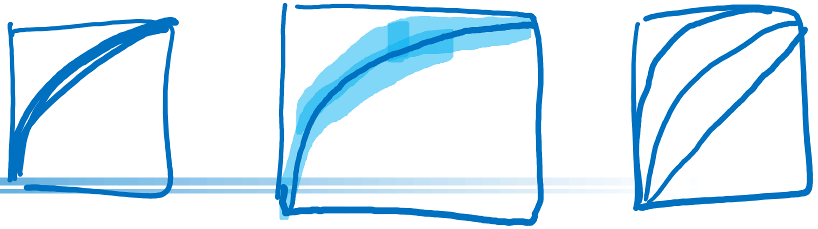
add β=0.99 to complete the ROC



ECE 580
Spring 2022
Stacy Tantum

T06.2: Binary Classifier Performance Evaluation

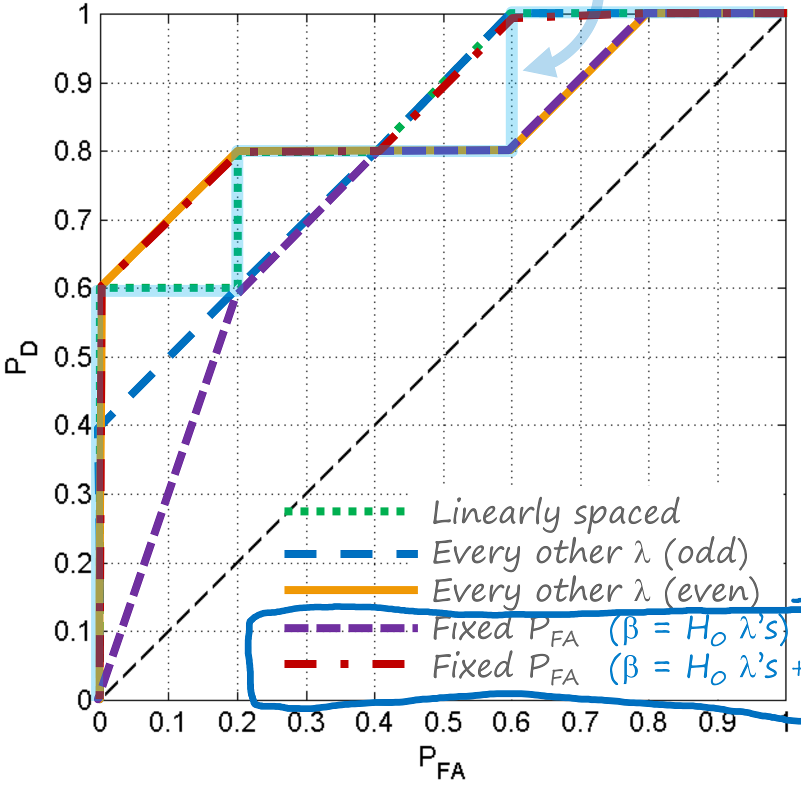
How Should We Select Decision Thresholds?



The ROC may vary a lot depending on how thresholds are selected

λ	truth
0	0
0.11	0
0.18	1
0.21	0
0.35	0
0.42	1
0.56	0
0.82	1
0.88	1
0.92	1

ROC for every λ is a threshold is a "stair step" ROC



Small # observations

→ amplifies differences

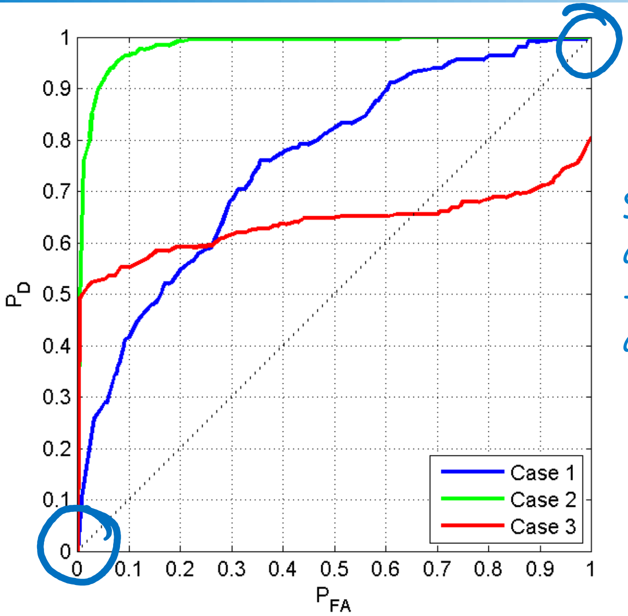
Large # observations

→ ROCs converge

Plotting ROCs

Plot the lines between the P_{FA}/P_D data points

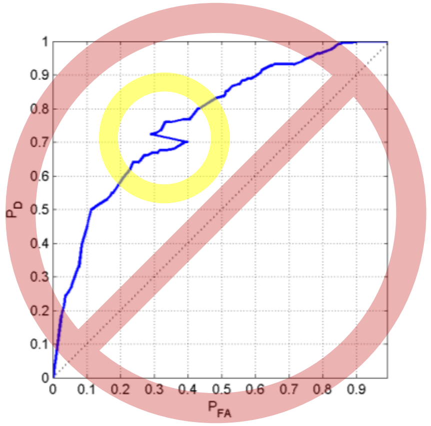
Symbols alone can be difficult to interpret if there are only a few P_{FA}/P_D points on the ROC

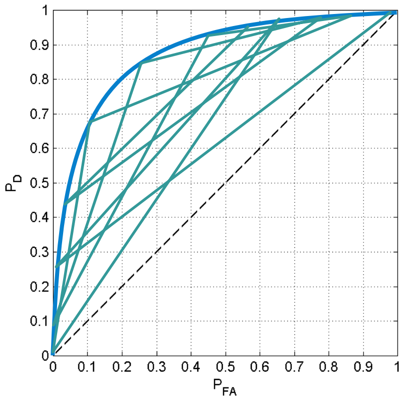


Some ROC functions assume $\lambda \in [0,1]$
→ Make sure you know what a function is assuming!

As threshold increases

- P_{FA} cannot increase
- P_D cannot increase
- If you see either P_{FA} or P_D increase when you increase the threshold, something is very wrong



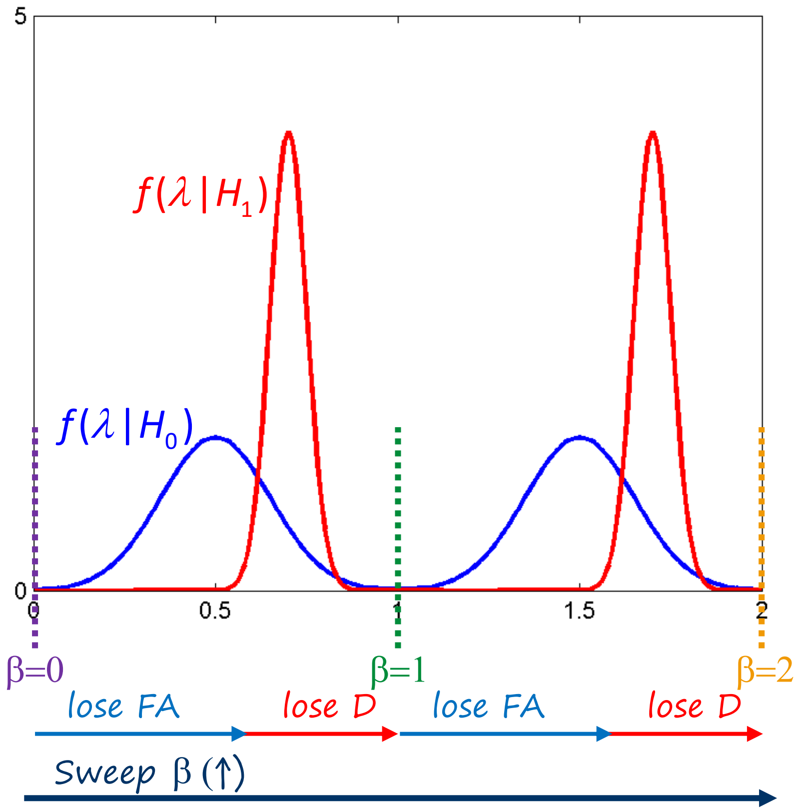


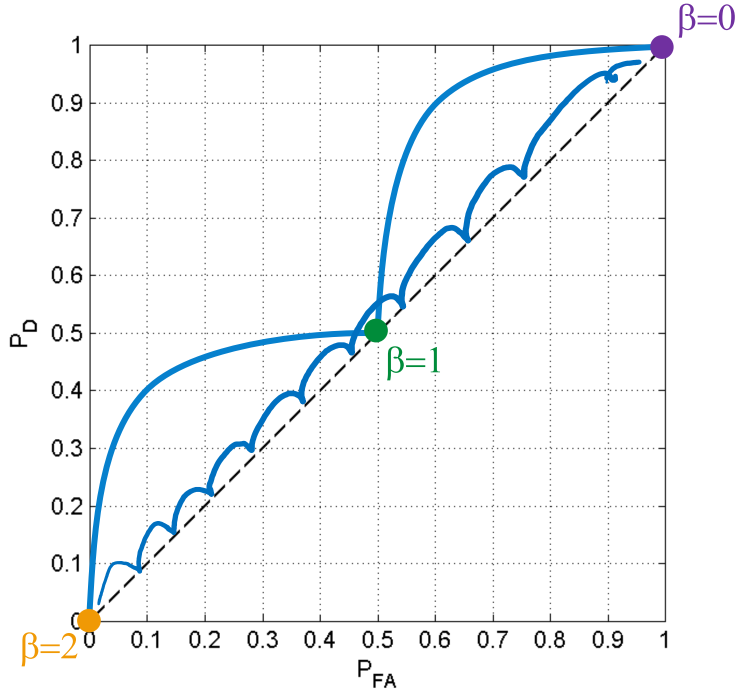
Ensure (P_{FA}, P_D) pairs are in order of increasing or decreasing threshold!

Decision Statistic pdfs & ROCs

Advice: λ distributions $\overset{\text{insight}}{\leftrightarrow}$ ROC shape

Common artifact of cross-validation when normalizing constants are not properly accounted for

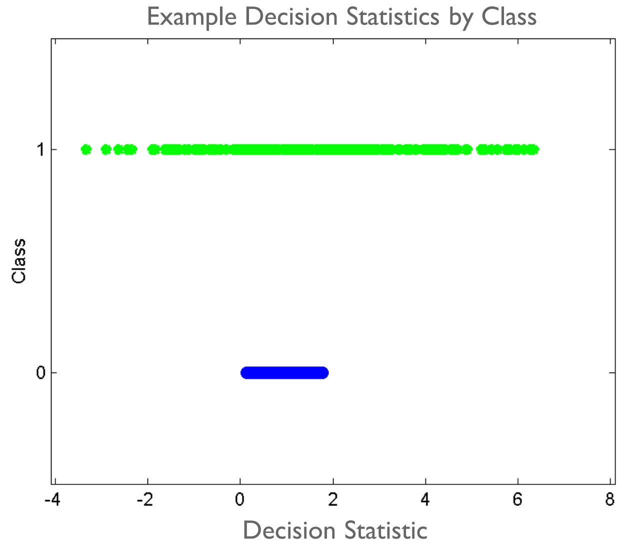




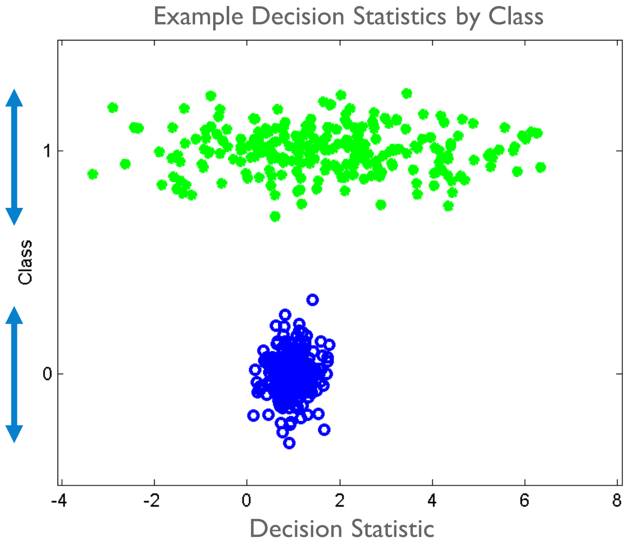
Visualizing Decision Statistics: Scatter Plots

Plot the decision statistic on the x-axis and class (target) on the y-axis

- Similar to plotting pdfs, but it may be easier to tell where individual decision statistics are falling in regions where there aren't many of them
- A few high H_0 decision statistics or low H_1 decision statistics can make the ROC look weird – this can help you understand that weirdness
- Can add a small amount of noise to the class variable to separate similar decision statistics



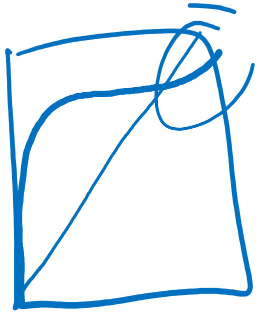
vertical coordinate is class plus small noise

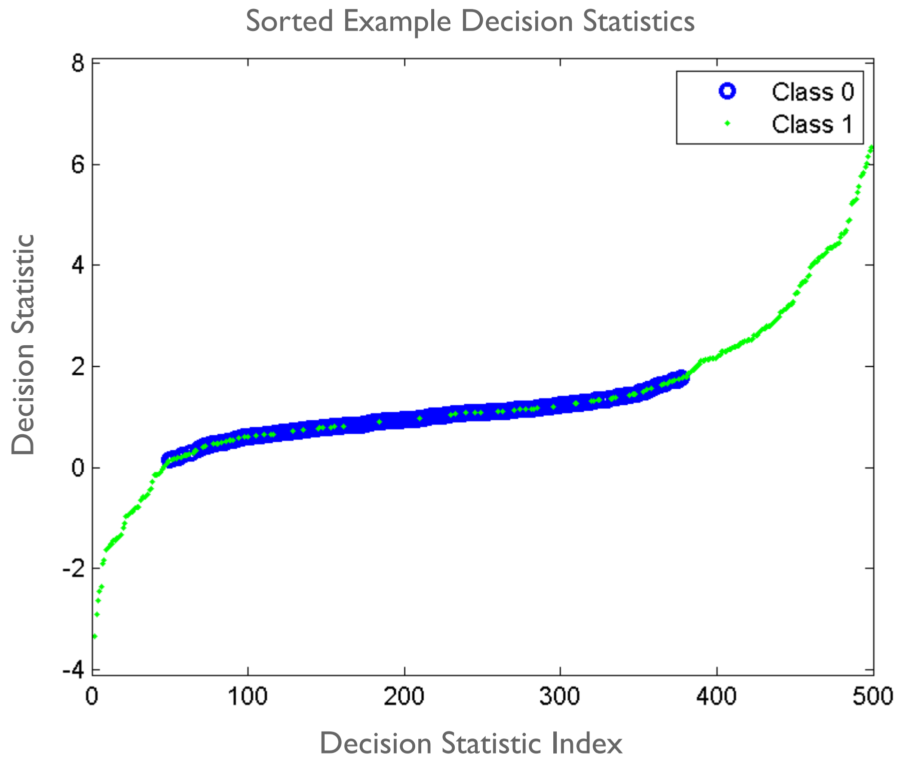


Visualizing Decision Statistics: Sorted Plots

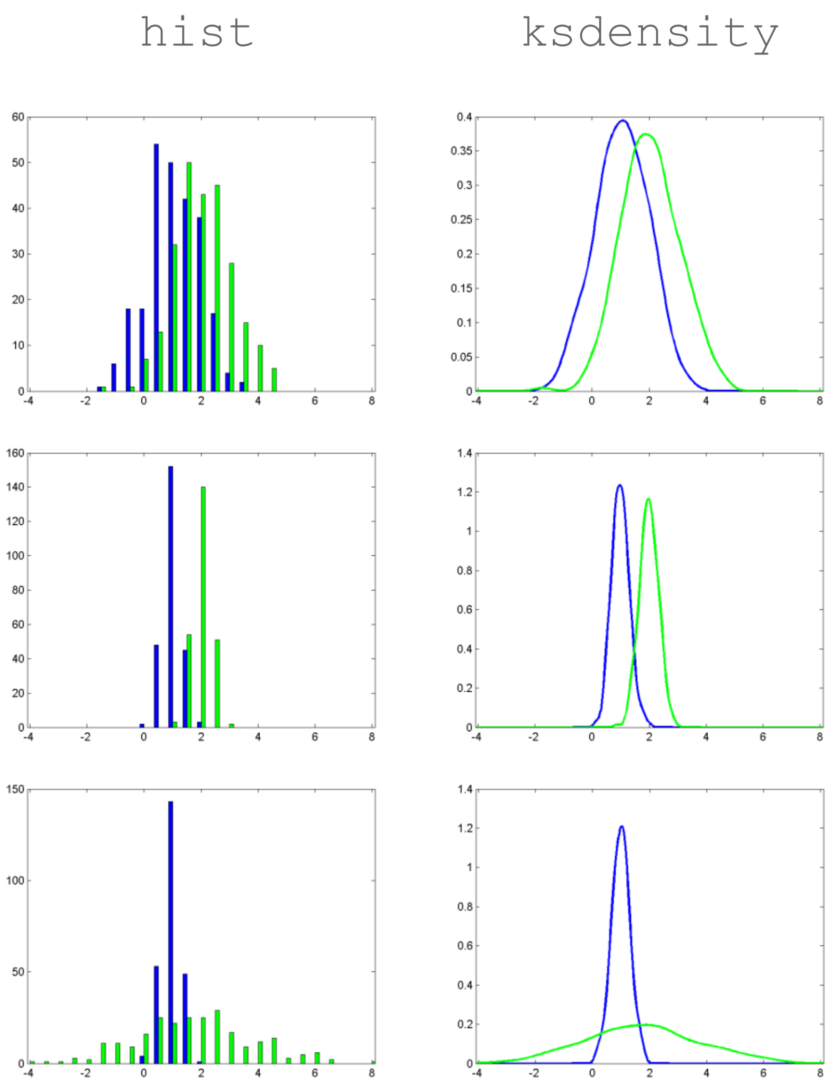
Sort the decision statistics from smallest to largest, and plot each one in a symbol that corresponds to its class

- Similar to plotting cdfs, but it may be easier to tell where individual decision statistics are falling in regions where there aren't many of them
- A few very high H_0 decision statistics or very low H_1 decision statistics can make the ROC look weird – this can help you understand that weirdness





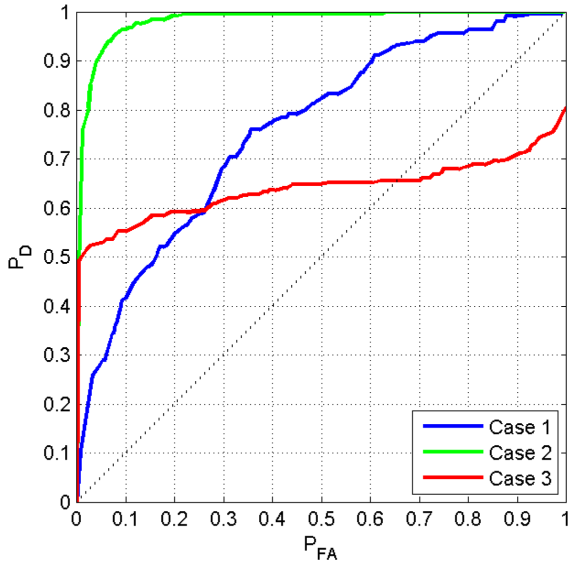
Visualizing Decision Statistics: pdfs



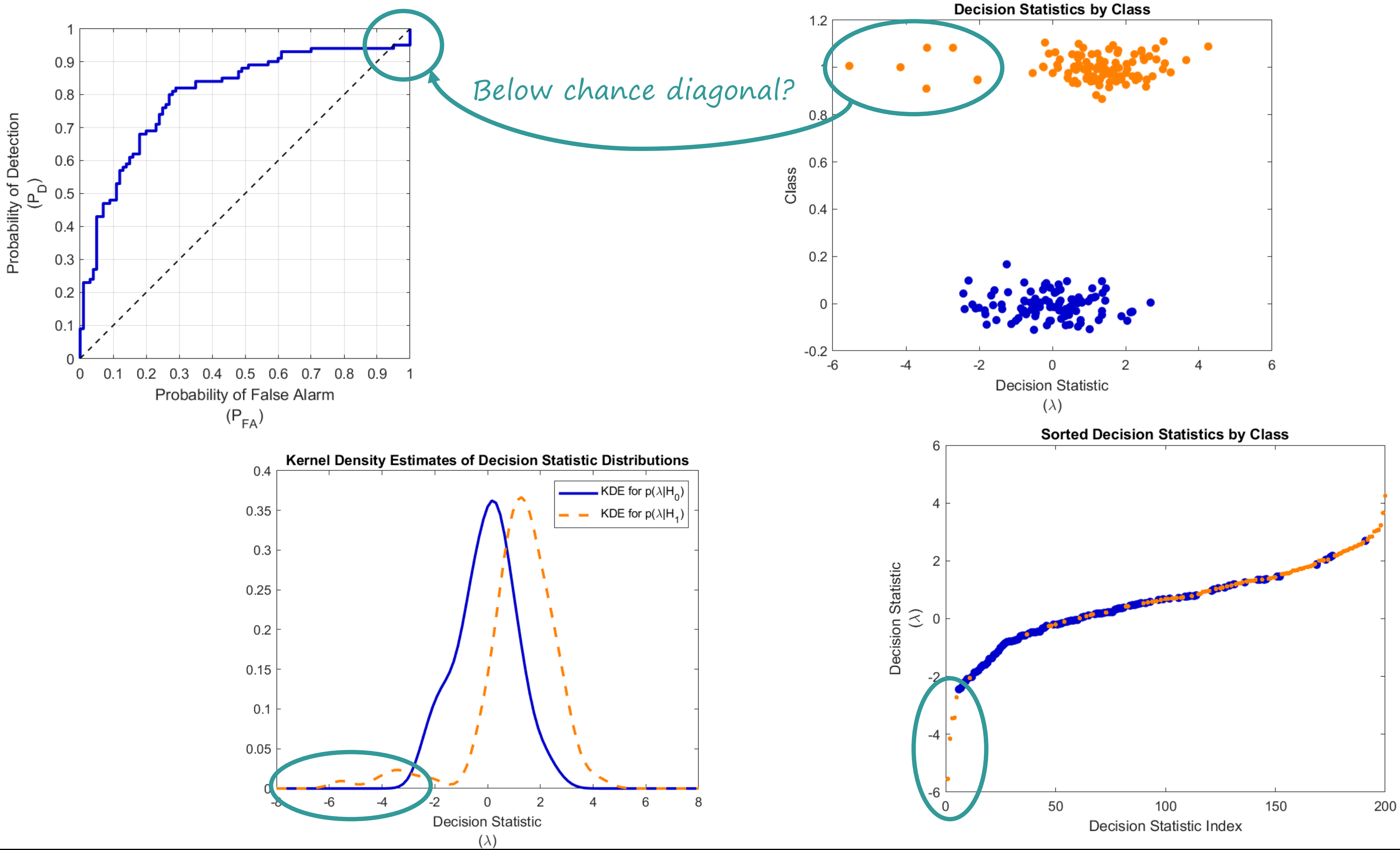
Visualize distributions
for each class separately

Smaller overlap between decision statistic pdfs:

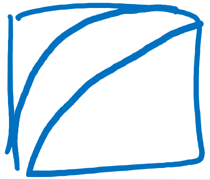
- Fewer opportunities for mis-classification in overlap region (better performance)
- Higher ROC
(closer to top-left corner of axes)



Visualizing Decision Statistics Can Help Make Sense of “Weird” ROCs

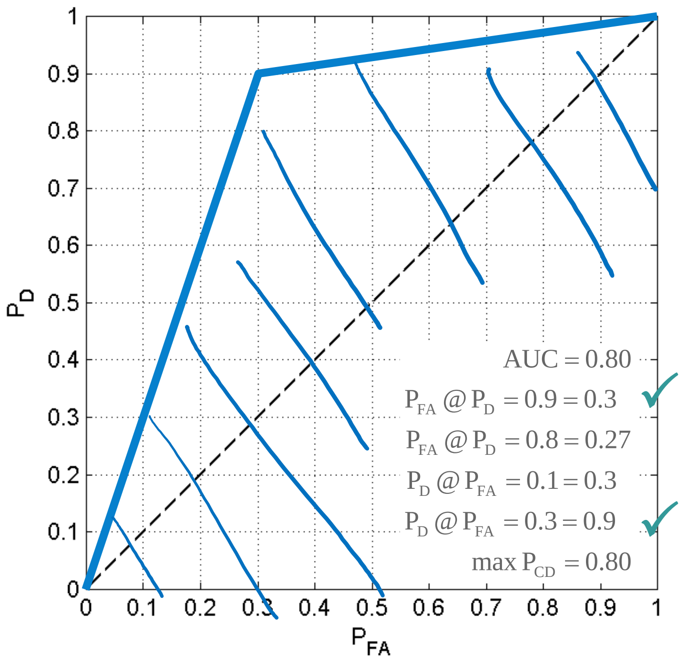


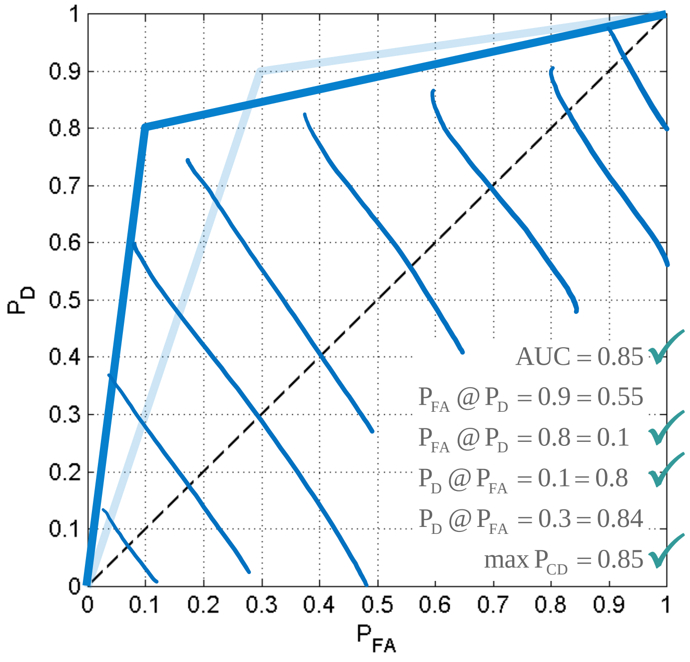
Comparing Classifier Performance

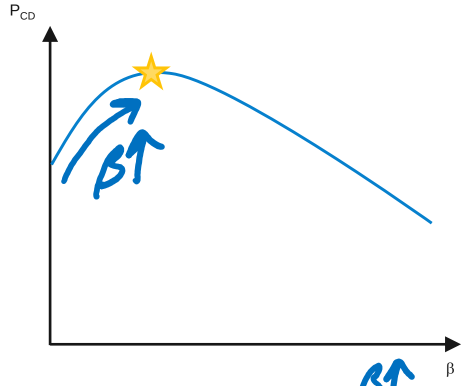


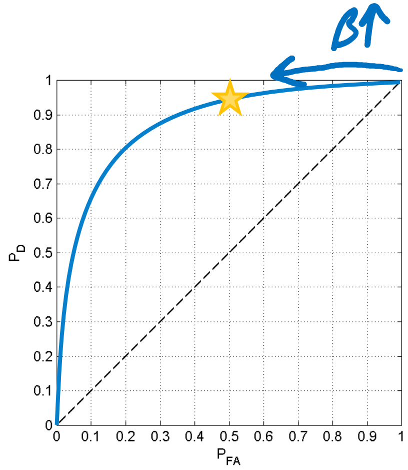
$p(\text{correct decision}) = P_{CD} = P_D p(H_1) + (1 - P_{FA})p(H_0)$ → depends on 1) priors $p(H_0)$ and $p(H_1)$

$= p(\text{choose } H_1 | H_1 \text{ true})p(H_1 \text{ true}) + (1 - p(\text{choose } H_1 | H_0 \text{ true}))p(H_0 \text{ true})$ 2) threshold β









Area Under Curve (AUC): higher → ROC more toward upper left

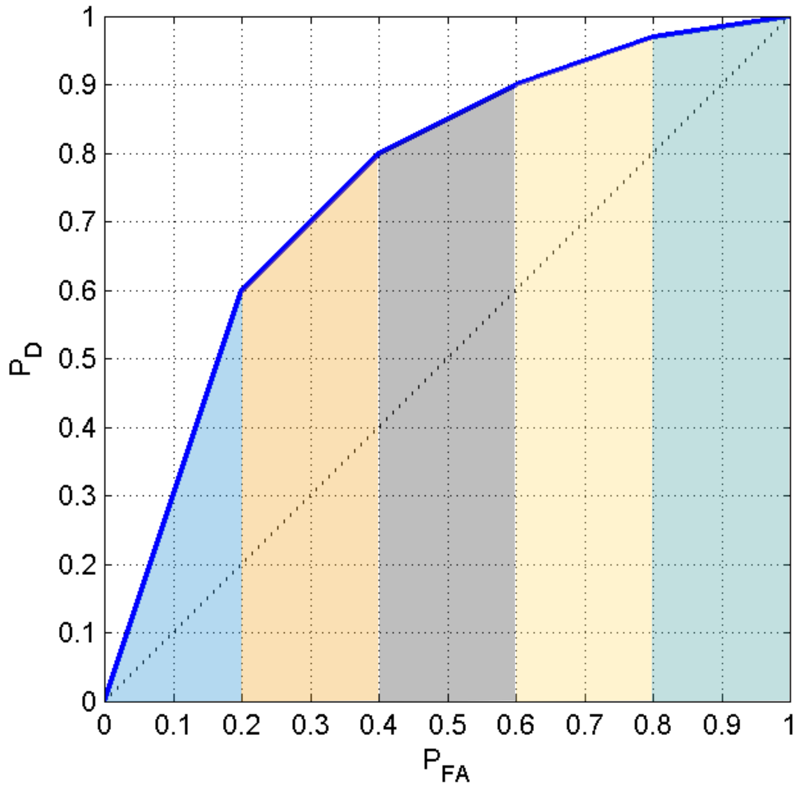
P_{FA}@P_D: lower → ROC more toward left @ fixed P_D

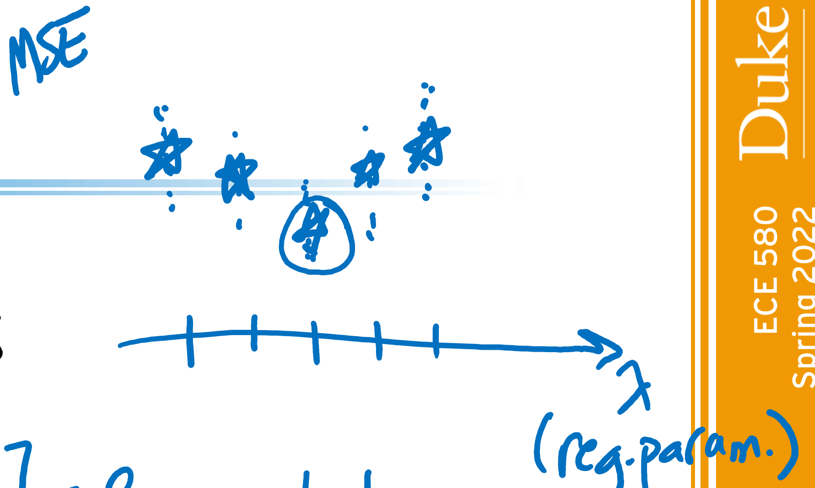
P_D@P_{FA}: higher → ROC more toward top @ fixed P_{FA}

Calculating AUC

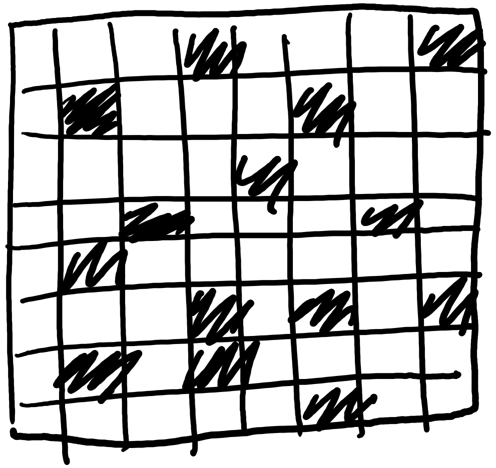
Trapezoidal integration (`trapz`) is a robust numerical method

- Provides an exact calculation for piece-wise linear curves – which is exactly what our ROC curve is
- Appropriate when variable of integration (P_{FA} or P_D) is not precisely evenly spaced
- Appropriate when the variable of integration is repeated, as for a “stair-step” ROC



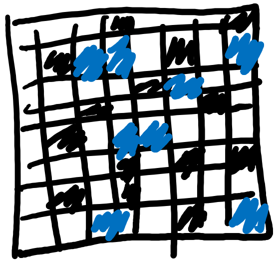


8×8 block



Fixed for all 20 iterations

1 iteration



■ $\lfloor \frac{S}{6} \rfloor = \lfloor \frac{50}{6} \rfloor = 8$ px testing

□ $S - \lfloor \frac{S}{6} \rfloor = 42$ px training

□ S=50 samples

■ 14 missing px

20 times: repeat process of randomly selecting 8 px for testing

(for each λ)