6.7930/HST.956 — Machine Learning for Healthcare
Spring 2025

Recitation 2

Bayes' Theorem
Differential Diagnosis
Eval Metrics

Ilker Demirel, 2/14/2025

About me

- 3rd year PhD student in MIT CSAIL
- Majored in Electrical & Electronics Engineering
- Got interested in ML4HC
    - Point-of-care ultrasound devices for non-expert use with RL
    - Personalized diabetes management with Bayesian Opt
- Currently working on causal inference + medical foundation models
    - How can we combine small/reliable data with large/biased data for better causal inference?
    - How can predict the future trajectory of a patient with autoregressive modeling

Pset submissions & MIMIC Access

- Minor updates
- Submit on gradescope.com
- Submit a single PDF file, append the code at the end
- Posted on Piazza as an announcement

Anyone does not have access to MIMIC
& Google Gemini on Vertex AI yet?

Bayes' Theorem

$$P(Y \mid X) = \frac{P(X \mid Y)P(Y)}{P(X)}$$

$$\propto P(X \mid Y)P(Y)$$

Updated Belief <— Data + Initial Belief

$$P(Y \mid X) \quad \propto \quad P(X \mid Y) \times P(Y)$$

Posterior prob.

Likelihood

Prior prob.

Bayes' Theorem

Updated Belief <— Data + Initial Belief

$$P(Y \mid X) = \frac{P(X \mid Y)P(Y)}{P(X)}$$

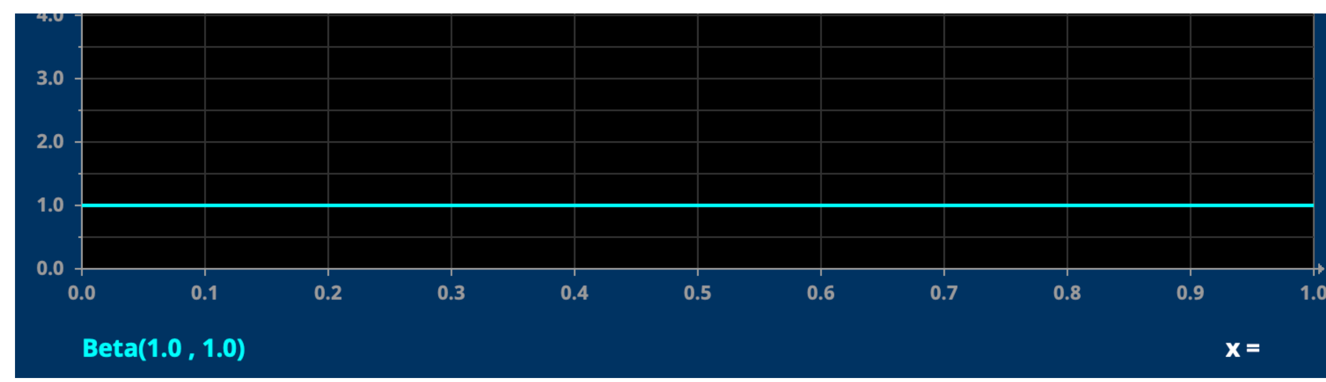
$$\propto P(X \mid Y)P(Y)$$

$P(Y \mid X) \quad \propto \quad P(X \mid Y) \times P(Y)$

Posterior prob. — *Likelihood* — *Prior* prob.

$Y:$ Probability of heads for an unfair coin

$P(Y):$



Bayes' Theorem

Updated Belief <— Data + Initial Belief

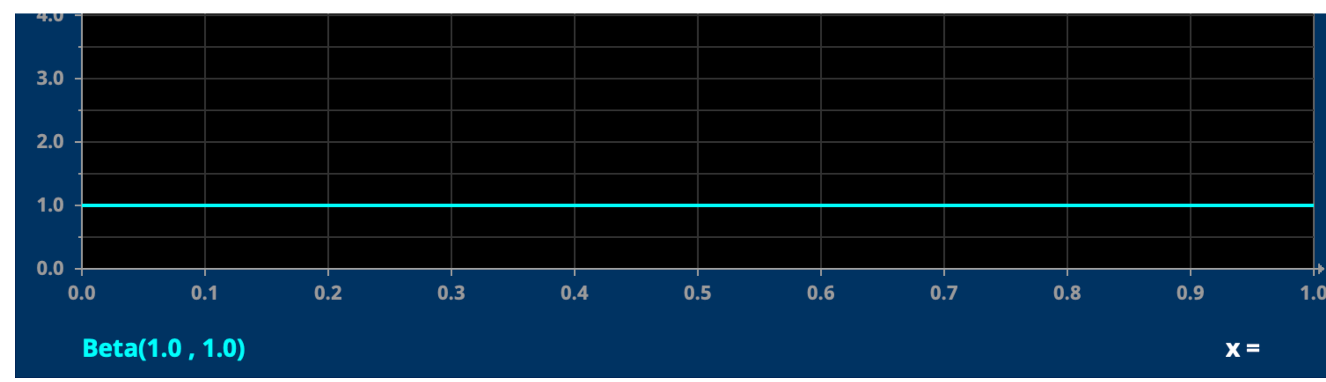
$$P(Y \mid X) = \frac{P(X \mid Y)P(Y)}{P(X)}$$
$$\propto P(X \mid Y)P(Y)$$

$$P(Y \mid X) \propto P(X \mid Y) \times P(Y)$$

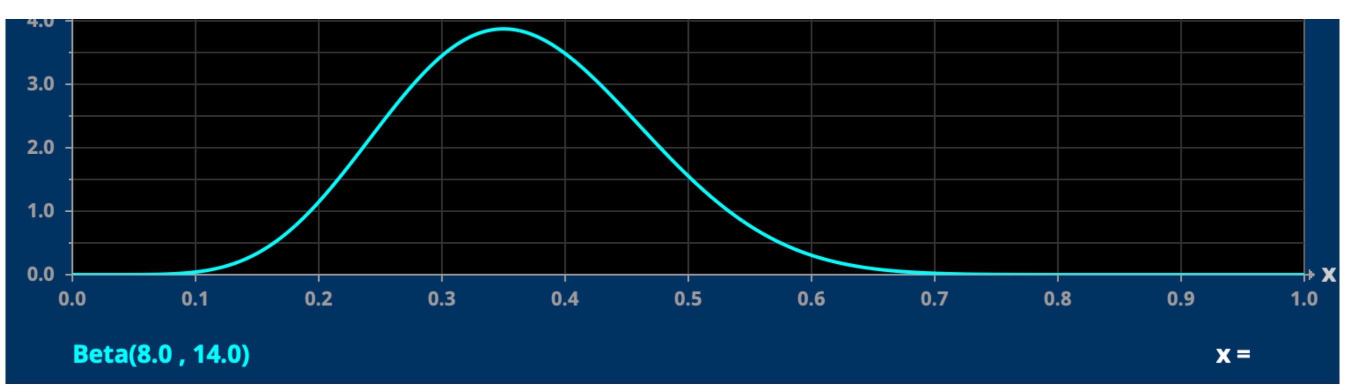
Posterior prob. — *Likelihood* — *Prior* prob.

$Y:$ Probability of heads for an unfair coin

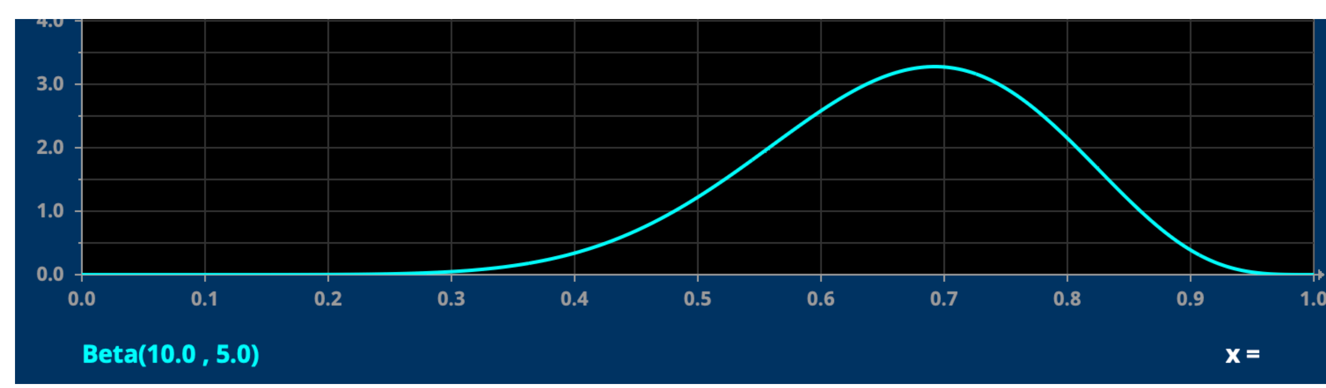
$P(Y):$



$P(Y \mid 7$ heads, 13 tails observed$)$



$P(Y \mid 9$ heads, 4 tails observed$)$



https://mathlets.org/mathlets/beta-distribution/

Bayes' Theorem

- ~1% of women aged 40-50 have breast cancer. $P(Y = 1) = 0.01$
- A woman with breast cancer has a 90% chance of a positive test from a mammogram. $P(T = 1 \mid Y = 1) = 0.9$
- A woman without has a 10% chance of a false positive result. $P(T = 1 \mid Y = 0) = 0.1$

Bayes' Theorem

- ~1% of women aged 40-50 have breast cancer. $P(Y=1)=0.01$
- A woman with breast cancer has a 90% chance of a positive test from a mammogram. $P(T=1 \mid Y=1)=0.9$
- A woman without has a 10% chance of a false positive result. $P(T=1 \mid Y=0)=0.1$

What is the probability a woman (aged 40-50) has breast cancer given that she just had a positive test?

Bayes' Theorem

- ~1% of women aged 40-50 have breast cancer. $P(Y=1)=0.01$
- A woman with breast cancer has a 90% chance of a positive test from a mammogram. $P(T=1 \mid Y=1)=0.9$
- A woman without has a 10% chance of a false positive result. $P(T=1 \mid Y=0)=0.1$

What is the probability a woman (aged 40-50) has breast cancer given that she just had a positive test?

$$P(Y=1 \mid T=1) = \frac{P(T=1 \mid Y=1)P(Y=1)}{P(T=1)}$$

$$= \frac{P(T=1 \mid Y=1)P(Y=1)}{P(T=1 \mid Y=1)P(Y=1) + P(T=1 \mid Y=0)P(Y=0)}$$

$$= \frac{0.9 \times 0.01}{0.9 \times 0.01 + 0.1 \times 0.99} = 0.083$$

Bayes' Theorem

According to the CDC: Compared to nonsmokers, women who smoke are about 13 times more likely to develop lung cancer.

If a woman is diagnosed with lung cancer, what is the probability that she is a smoker?

Bayes' Theorem

According to the CDC: Compared to nonsmokers, women who smoke are about 13 times more likely to develop lung cancer.

If a woman is diagnosed with lung cancer, what is the probability that she is a smoker?

$$
\begin{aligned}
P(S=1 \mid L=1) &= \frac{P(L=1 \mid S=1)P(S=1)}{P(L=1)} \\
&= \frac{P(L=1 \mid S=1)P(S=1)}{P(L=1 \mid S=1)P(S=1) + P(L=1 \mid S=0)P(S=0)} \\
&= \frac{13x \times y}{13xy \times x(1-y)} \\
&= \frac{13y}{13y + (1-y)}
\end{aligned}
$$

Bayes' Theorem

According to the CDC: Compared to nonsmokers, women who smoke are about 13 times more likely to develop lung cancer.

If a woman is diagnosed with lung cancer, what is the probability that she is a smoker?

$$P(S=1 \mid L=1) = \frac{P(L=1 \mid S=1)P(S=1)}{P(L=1)}$$

$$= \frac{P(L=1 \mid S=1)P(S=1)}{P(L=1 \mid S=1)P(S=1) + P(L=1 \mid S=0)P(S=0)}$$

$$= \frac{13x \times y}{13xy \times x(1-y)}$$

$$= \frac{13y}{13y + (1-y)}$$

$$= 0.74$$

The fraction of women who smoke is y=17.9% according to CDC (2009).

Cannot compute without the prior

Naive Bayes

Conditioning on Multiple Variables

$$
\begin{aligned}
P(Y \mid X_1, X_2, X_3) &\propto P(X_1, X_2, X_3 \mid Y)P(Y) \\
&= P(X_1 \mid Y)P(X_2 \mid X_1, Y)P(X_3 \mid X_1, X_2, Y)P(Y)
\end{aligned}
$$

Naive Bayes

Conditioning on Multiple Variables

$$P(Y \mid X_1, X_2, X_3) \propto P(X_1, X_2, X_3 \mid Y)P(Y)$$
$$= P(X_1 \mid Y)P(X_2 \mid X_1, Y)P(X_3 \mid X_1, X_2, Y)P(Y)$$

Quickly becomes impractical to estimate

Naive Bayes

Multiple Symptoms/Tests (Naive Bayes)

$$P(Y \mid X_1, X_2, X_3) \propto P(X_1, X_2, X_3 \mid Y)P(Y)$$

$$= P(X_1 \mid Y)P(X_2 \mid X_1, Y)P(X_3 \mid X_1, X_2, Y)P(Y)$$

Conditioned on Y, X_i are independent →

$$= P(X_1 \mid Y)P(X_2 \mid Y)P(X_3 \mid Y)P(Y)$$

Naive Bayes

Conditioning on Multiple Variables

$$P(Y \mid X_1, X_2, X_3) \propto P(X_1, X_2, X_3 \mid Y)P(Y)$$

$$= P(X_1 \mid Y)P(X_2 \mid X_1, Y)P(X_3 \mid X_1, X_2, Y)P(Y)$$

Naive Bayes ⟶ $= P(X_1 \mid Y)P(X_2 \mid Y)P(X_3 \mid Y)P(Y)$

Conditioned on Y, X_i are independent

Useful when new evidence arrives sequentially (e.g., new lab results)

Posterior becomes prior, keep updating as before

More nuanced & realistic conditional independencies can be baked in depending on domain knowledge

Naive Bayes

- Patient presents in the clinic, and the primary care physician writes down the following note:
    - "25 yo, has SOB, chronic coughing, active smoker 1ppd"

$X_1 = 1$ if > 65yo	$X_1 = 0$ otherwise
$X_2 = 1$ if SOB	$X_2 = 0$ otherwise
$X_3 = 1$ if cough	$X_3 = 0$ otherwise
$X_4 = 1$ if smoker	$X_4 = 0$ otherwise

$(X_1, X_2, X_3, X_4) = (0, 1, 1, 1)$

Y : Lung cancer

$P(Y = 1) \approx 0.0007$

Naive Bayes

- Patient presents in the clinic, and the primary care physician writes down the following note:
  - "25 yo, has SOB, chronic coughing, active smoker 1ppd"

$X_1 = 1$	if > 65yo	$X_1 = 0$	otherwise
$X_2 = 1$	if SOB	$X_2 = 0$	otherwise
$X_3 = 1$	if cough	$X_3 = 0$	otherwise
$X_4 = 1$	if smoker	$X_4 = 0$	otherwise

(X_1, X_2, X_3, X_4) = (0, 1, 1, 1)

Y : Lung cancer

$P(Y=1) \approx 0.0007$

$$P(Y=1 \mid X_1=0, X_2=1, X_3=1, X_4=1)$$
$$\propto P(X_1=0 \mid Y=1)$$
$$\times P(X_2=1 \mid Y=1)$$
$$\times P(X_3=1 \mid Y=1)$$
$$\times P(X_4=1 \mid Y=1)$$
$$\times P(Y=1)$$

Estimate everything (also for $Y=0$) & make decisions

Evaluation Metrics

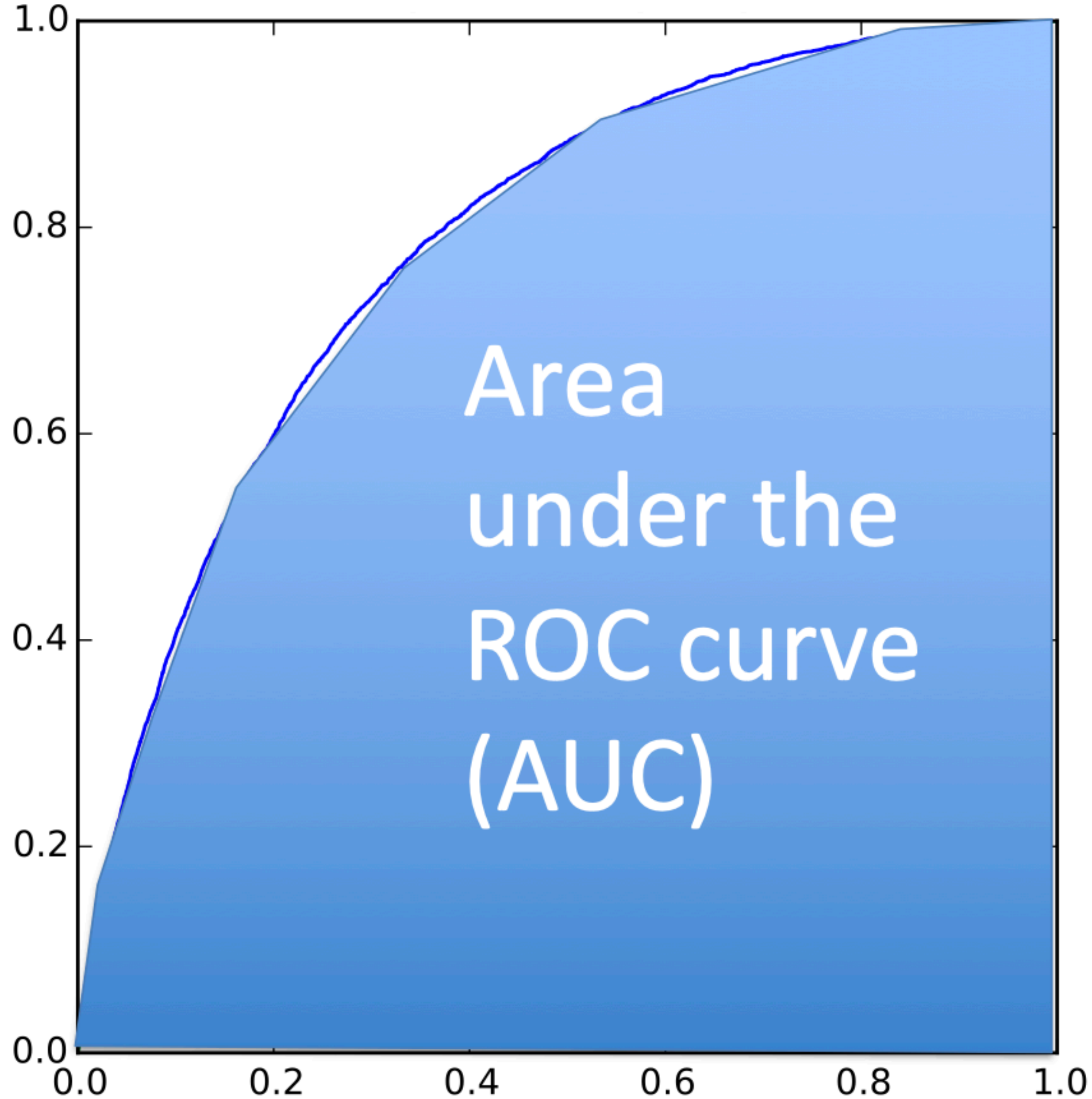
Sources: [1][2][3][4][5][6][7][8] view · talk · edit

		Predicted condition			
	Total population $= P + N$	**Predicted positive**	**Predicted negative**	Informedness, bookmaker informedness (BM) $= TPR + TNR - 1$	Prevalence threshold (PT) $= \frac{\sqrt{TPR \times FPR} - FPR}{TPR - FPR}$
Actual condition	**Positive (P)** [a]	**True positive** (TP), hit[b]	**False negative** (FN), miss, underestimation	True positive rate (TPR), recall, sensitivity (SEN), probability of detection, hit rate, power $= \frac{TP}{P} = 1 - FNR$	False negative rate (FNR), miss rate type II error [c] $= \frac{FN}{P} = 1 - TPR$
	Negative (N) [d]	**False positive** (FP), false alarm, overestimation	**True negative** (TN), correct rejection[e]	False positive rate (FPR), probability of false alarm, fall-out type I error [f] $= \frac{FP}{N} = 1 - TNR$	True negative rate (TNR), specificity (SPC), selectivity $= \frac{TN}{N} = 1 - FPR$
	Prevalence $= \frac{P}{P + N}$	Positive predictive value (PPV), precision $= \frac{TP}{TP + FP} = 1 - FDR$	False omission rate (FOR) $= \frac{FN}{TN + FN} = 1 - NPV$	Positive likelihood ratio (LR+) $= \frac{TPR}{FPR}$	Negative likelihood ratio (LR–) $= \frac{FNR}{TNR}$
	Accuracy (ACC) $= \frac{TP + TN}{P + N}$	False discovery rate (FDR) $= \frac{FP}{TP + FP} = 1 - PPV$	Negative predictive value (NPV) $= \frac{TN}{TN + FN} = 1 - FOR$	Markedness (MK), deltaP (Δp) $= PPV + NPV - 1$	Diagnostic odds ratio (DOR) $= \frac{LR+}{LR-}$
	Balanced accuracy (BA) $= \frac{TPR + TNR}{2}$	F_1 score $= \frac{2\, PPV \times TPR}{PPV + TPR} = \frac{2\, TP}{2\, TP + FP + FN}$	Fowlkes–Mallows index (FM) $= \sqrt{PPV \times TPR}$	Matthews correlation coefficient (MCC) $= \sqrt{TPR \times TNR \times PPV \times NPV} - \sqrt{FNR \times FPR \times FOR \times FDR}$	Threat score (TS), critical success index (CSI), Jaccard index $= \frac{TP}{TP + FN + FP}$

https://en.wikipedia.org/wiki/Evaluation_of_binary_classifiers

Receiver-operator characteristic curve

True positive rate (sensitivity) … i.e. TP/P



False positive rate (1-specificity)… i.e. FP/N

Diabetes 1-year gap

AUC = Probability that algorithm ranks a positive patient over a negative patient

Invariant to amount of class imbalance

How does one get the AUC curve?

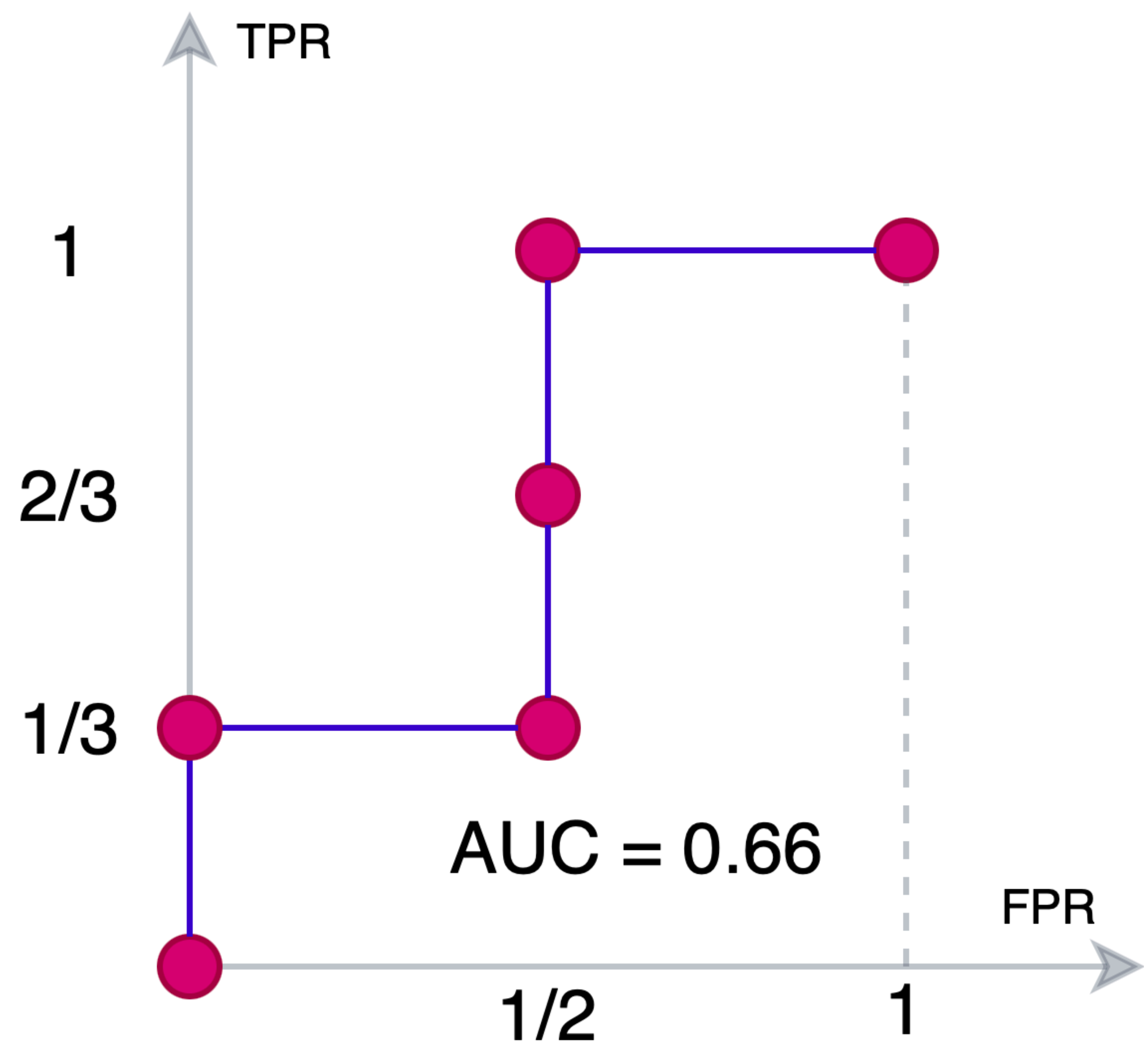
- $Y = (1,\ 0,\ 1,\ 1,\ 0)$
- $\hat{P}(Y = 1) = (0.9,\ 0.1,\ 0.4,\ 0.6,\ 0.8)$

How does one get the AUC curve?

- $Y = (1,\ 0,\ 1,\ 1,\ 0)$
- $\hat{P}(Y = 1) = (0.9,\ 0.1,\ 0.4,\ 0.6,\ 0.8)$
- Threshold: 0 —> $\hat{Y} = (1,\ 1,\ 1,\ 1,\ 1)$ —> TPR = 3/3, FPR = 2/2
- Threshold: 0.2 —> $\hat{Y} = (1,\ 0,\ 1,\ 1,\ 1)$ —> TPR = 3/3, FPR = 1/2
- Threshold: 0.5 —> $\hat{Y} = (1,\ 0,\ 0,\ 1,\ 1)$ —> TPR = 2/3, FPR = 1/2
- Threshold: 0.7 —> $\hat{Y} = (1,\ 0,\ 0,\ 0,\ 1)$ —> TPR = 1/3, FPR = 1/2
- Threshold: 0.85 —> $\hat{Y} = (1,\ 0,\ 0,\ 0,\ 0)$ —> TPR = 1/3, FPR = 0/2
- Threshold: 0.95 —> $\hat{Y} = (0,\ 0,\ 0,\ 0,\ 0)$ —> TPR = 0/3, FPR = 0/2

How does one get the AUC curve?

- $Y = (1,\ 0,\ 1,\ 1,\ 0)$
- $\hat{P}(Y = 1) = (0.9,\ 0.1,\ 0.4,\ 0.6,\ 0.8)$



Be careful about AUROC

	Pred Pos	Pred Neg
True Pos	100	0
True Neg	200	1800

$$\text{TPR} = \frac{100}{100 + 0} = 1$$

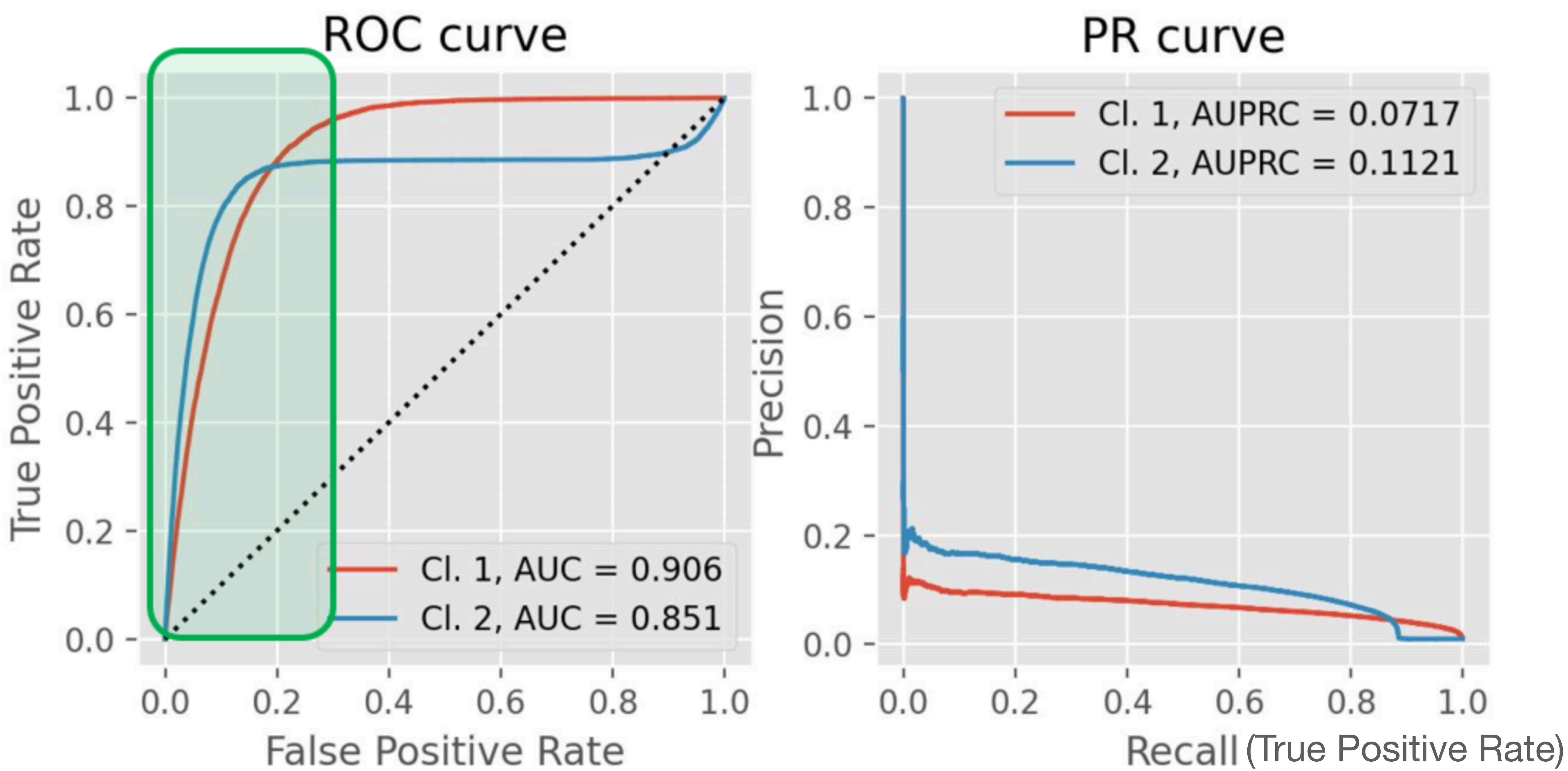
$$\text{FPR} = \frac{200}{200 + 1800} = 0.1$$

$$\text{PPV} = \frac{100}{100 + 200} = 0.33$$

Good AUROC

Bad AUPRC

Check AUPRC too



https://medium.com/towards-data-science/demystifying-roc-and-precision-recall-curves-d30f3fad2cbf

https://stats.stackexchange.com/questions/251175/what-is-baseline-in-precision-recall-curve