Lab6: Logistic Regression and Metrics

Department of Computer Science,
National Tsing Hua University, Taiwan

2020.10.08

Outline

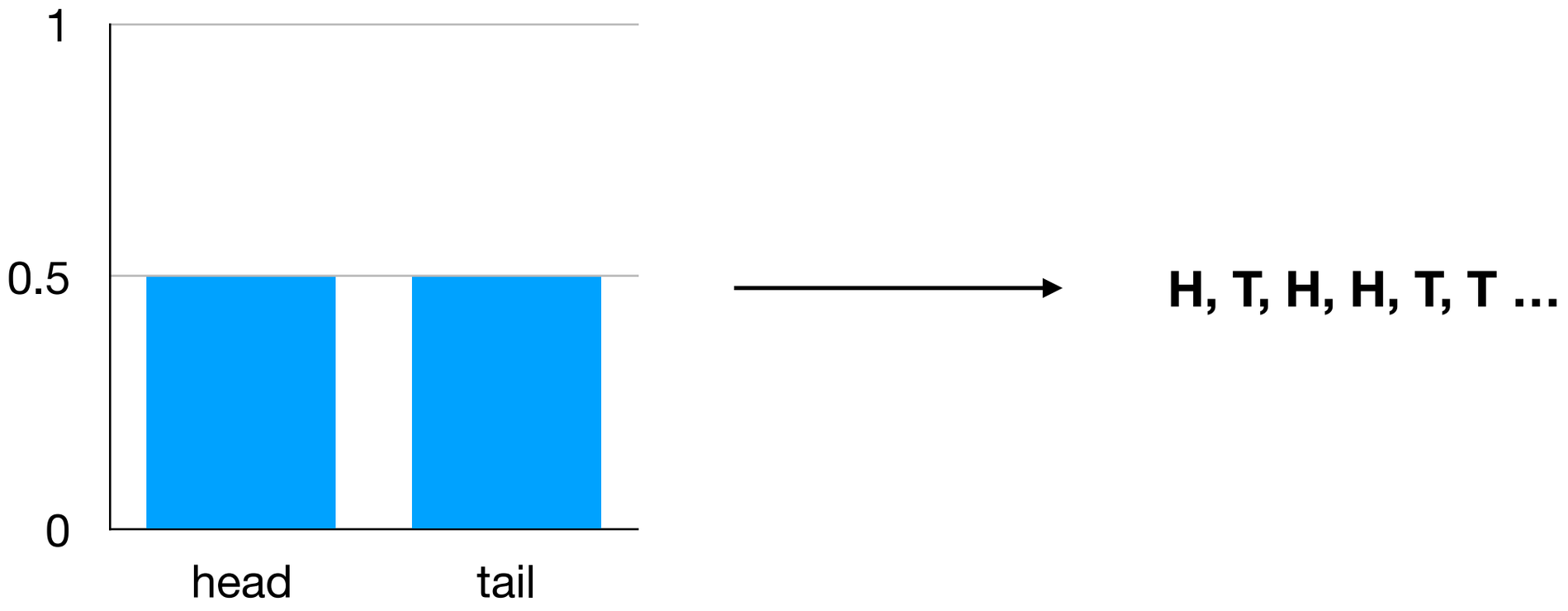
- Brief Review: Logistic Regression
  - Maximum likelihood in Logistic Regression
  - Implementation
- Common Evaluation Metrics for Binary Classification
  - Confusion Matrix
  - Soft Classifiers - ROC Curve

Outline

- Brief Review: Logistic Regression
  - Maximum likelihood in Logistic Regression
  - Implementation
- Common Evaluation Metrics for Binary Classification
  - Confusion Matrix
  - Soft Classifiers - ROC Curve

Maximum Likelihood

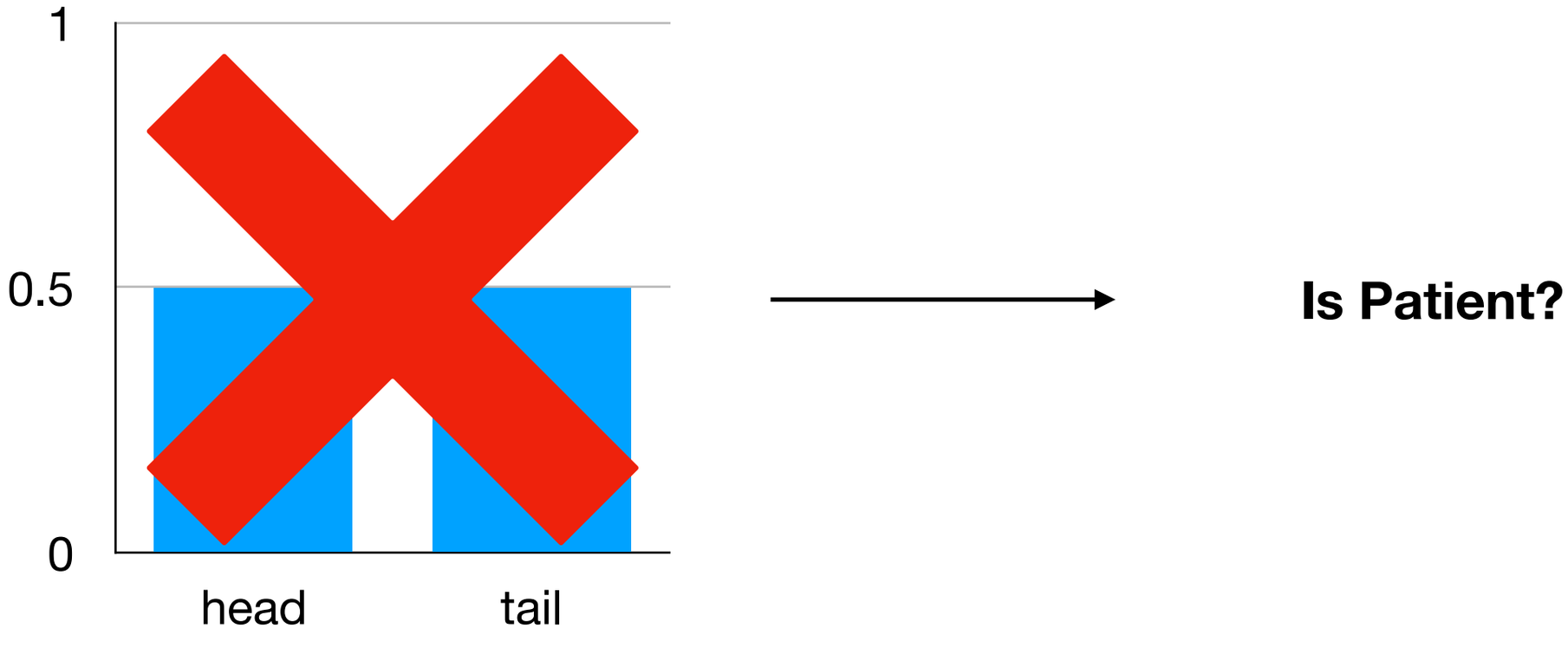
- Flipping coin: we have already know ground truth distribution. For example, $\mathrm{P}(x = head) = 1/2$ and $\mathrm{P}(x = tail) = 1/2$.



Maximum Likelihood

- Flipping coin: we have already know ground truth distribution. For example, $\mathrm{P}(x = head) = 1/2$ and $\mathrm{P}(x = tail) = 1/2$.

- However, in many tasks, the ground truth distributions are never known, e.g., probability distribution of getting COVID-19.



Maximum Likelihood

- The process to approximate the distribution:

  - First, we assume the proportion of people diagnosed with a disease follows Binomial distribution, e.g., $X \sim Bin(A,\ \rho)$.

Maximum Likelihood

- The process to approximate the distribution:

  - First, we assume the proportion of people diagnosed with a disease follows Binomial distribution, e.g., $X \sim Bin(A,\ \rho)$.
    Where A is number of person diagnosed, ρ is illness rate.

Maximum Likelihood

- The process to approximate the distribution:

  - First, we assume the proportion of people diagnosed with a disease follows Binomial distribution, e.g., $X \sim Bin(A,\ \rho)$. Where A is number of person diagnosed, ρ is illness rate.

  - If there are 4 patients out of 10 people, the number of Bernoulli trials would be 10, i.e., $X \sim Bin(10,\ \rho)$

Maximum Likelihood

- The process to approximate the distribution:

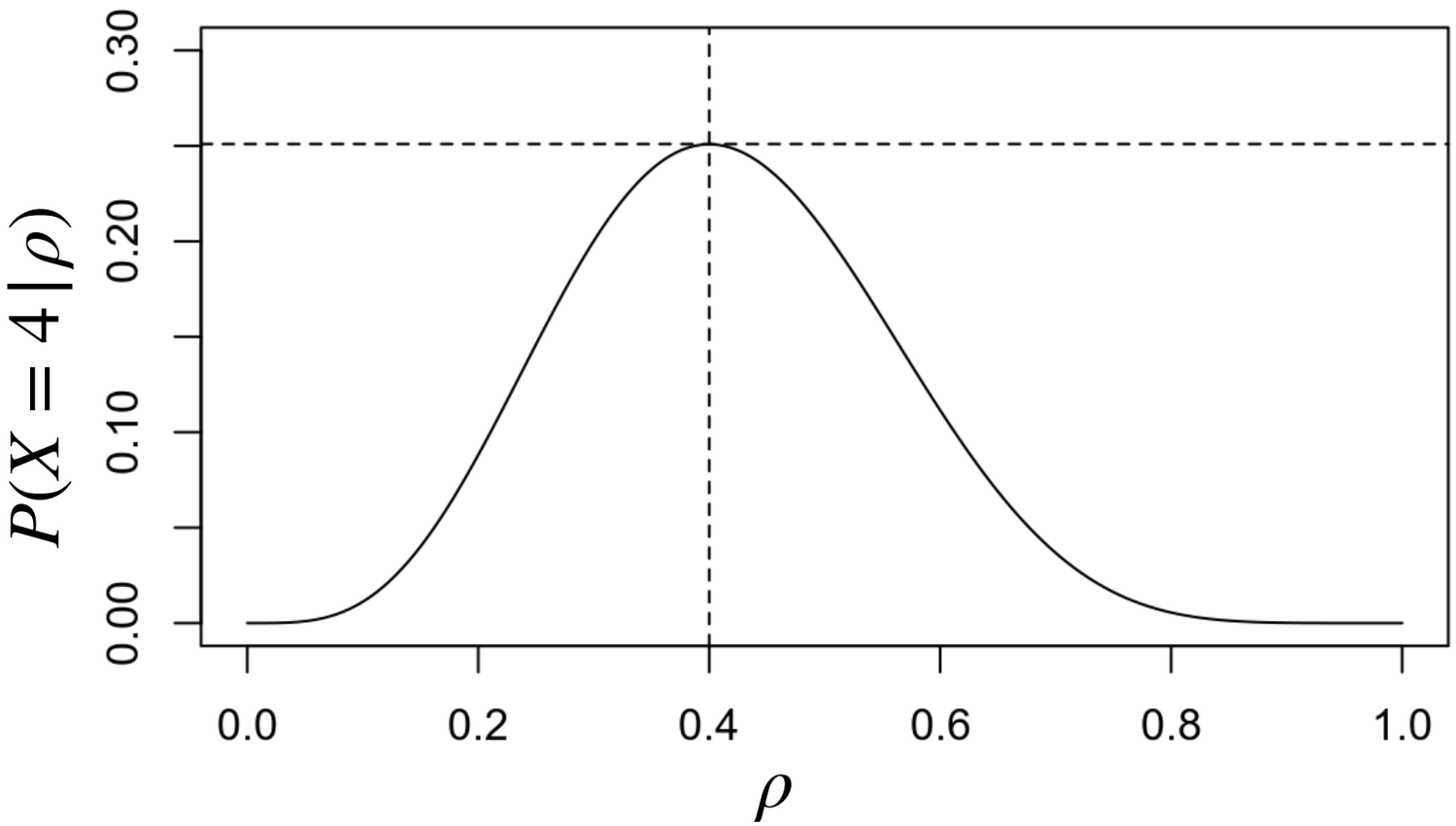
  - First, we assume the proportion of people diagnosed with a disease follows Binomial distribution, e.g., $X \sim Bin(A,\ \rho)$. Where A is number of person diagnosed, ρ is illness rate.

  - If there are 4 patients out of 10 people, the number of Bernoulli trials would be 10, i.e., $X \sim Bin(10,\ \rho)$

$$\mathrm{P}(X = 4 \mid \rho) = C_4^{10} \rho^4 (1-\rho)^{(10-4)}$$

Maximum Likelihood

$$\mathrm{P}(X = 4 \,|\, \rho) = C_4^{10} \rho^4 (1 - \rho)^{(10-4)}$$



Outline

- Brief Review: Logistic Regression
  - Maximum likelihood in Logistic Regression
  - Implementation
- Common Evaluation Metrics for Binary Classification
  - Confusion Matrix
  - Soft Classifiers - ROC Curve

Logistic Regression

- In logistic regression, we solve maximum log-likelihood instead.

$$\arg\max_{\boldsymbol{w}} \log \mathrm{P}(\mathbb{X} \mid \boldsymbol{w})$$

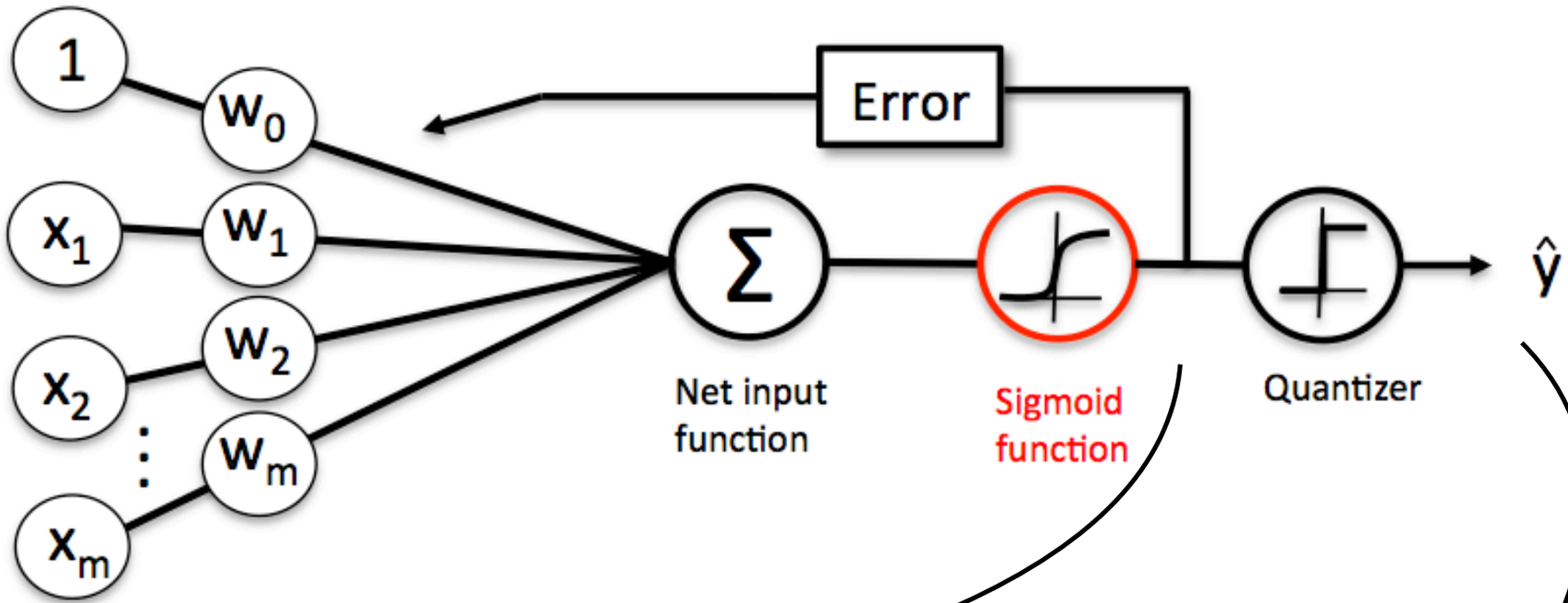
- Update with gradient decent:

$$\boldsymbol{w}^{(t+1)} = \boldsymbol{w}^{(t)} - \eta \nabla_{\boldsymbol{w}} \log \mathrm{P}(\mathbb{X} \mid \boldsymbol{w}^{(t)})$$

where

$$\nabla_{\boldsymbol{w}} \log \mathrm{P}(\mathbb{X} \mid \boldsymbol{w}^{(t)}) = \sum_{t=1}^{N} [y'^{(i)} - \sigma(\boldsymbol{w}^{(t)\top} \boldsymbol{x}^{(i)})] \boldsymbol{x}^{(i)}, \quad y' = \frac{y+1}{2}$$

Logistic Regression



$$P(y \mid \boldsymbol{x}; \boldsymbol{w}) = \sigma(\boldsymbol{w}^\top \boldsymbol{x})^{y'}[1 - \sigma(\boldsymbol{w}^\top \boldsymbol{x})]^{(1-y')}$$

Soft prediction

$$\arg\max_{y} \{\sigma(\boldsymbol{w}^\top \boldsymbol{x}), 1 - \sigma(\boldsymbol{w}^\top \boldsymbol{x})\} = \text{sign}(\boldsymbol{w}^\top \boldsymbol{x})$$

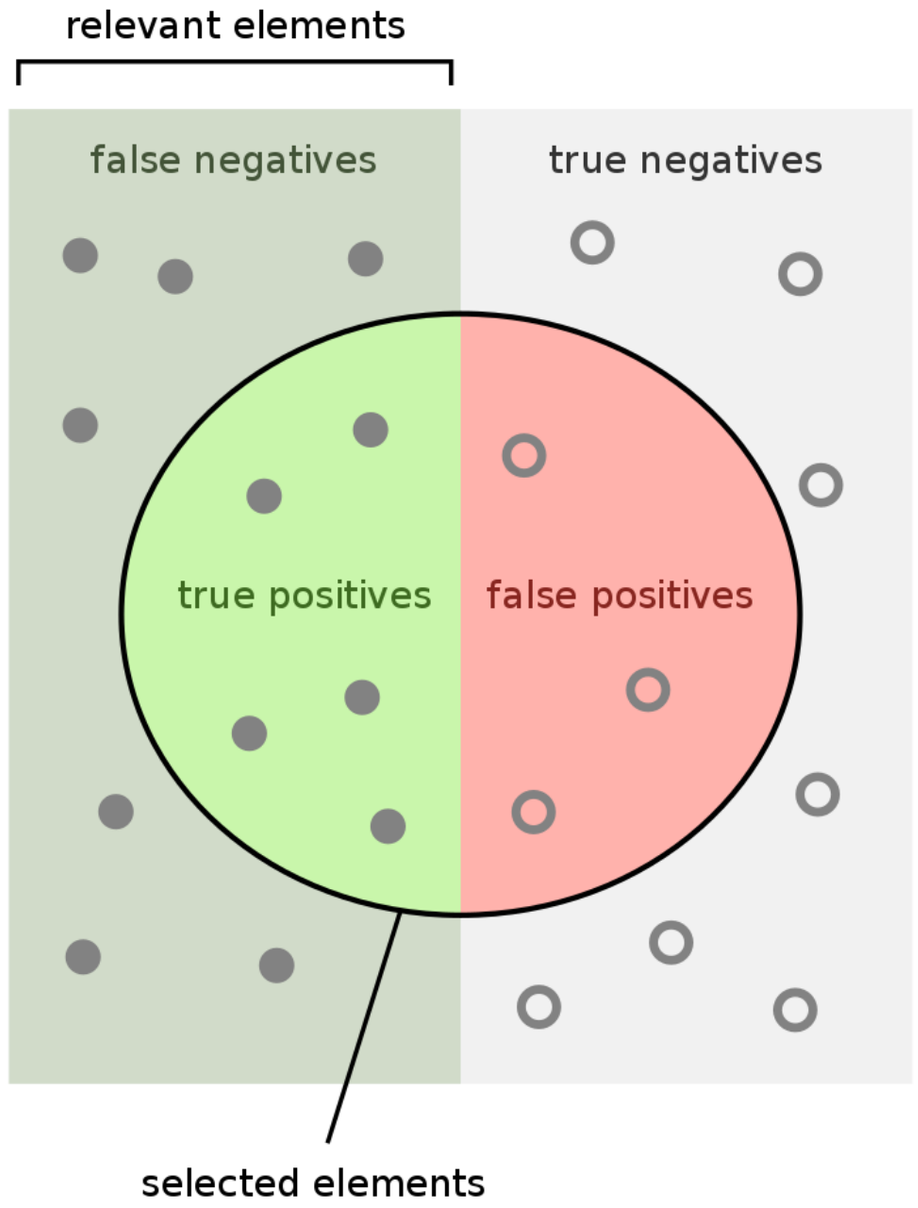
Label prediction

Outline

- Brief Review: Logistic Regression
  - Maximum likelihood in Logistic Regression
  - Implementation
- Common Evaluation Metrics for Binary Classification
  - Confusion Matrix
  - Soft Classifiers - ROC Curve

Confusion Matrix

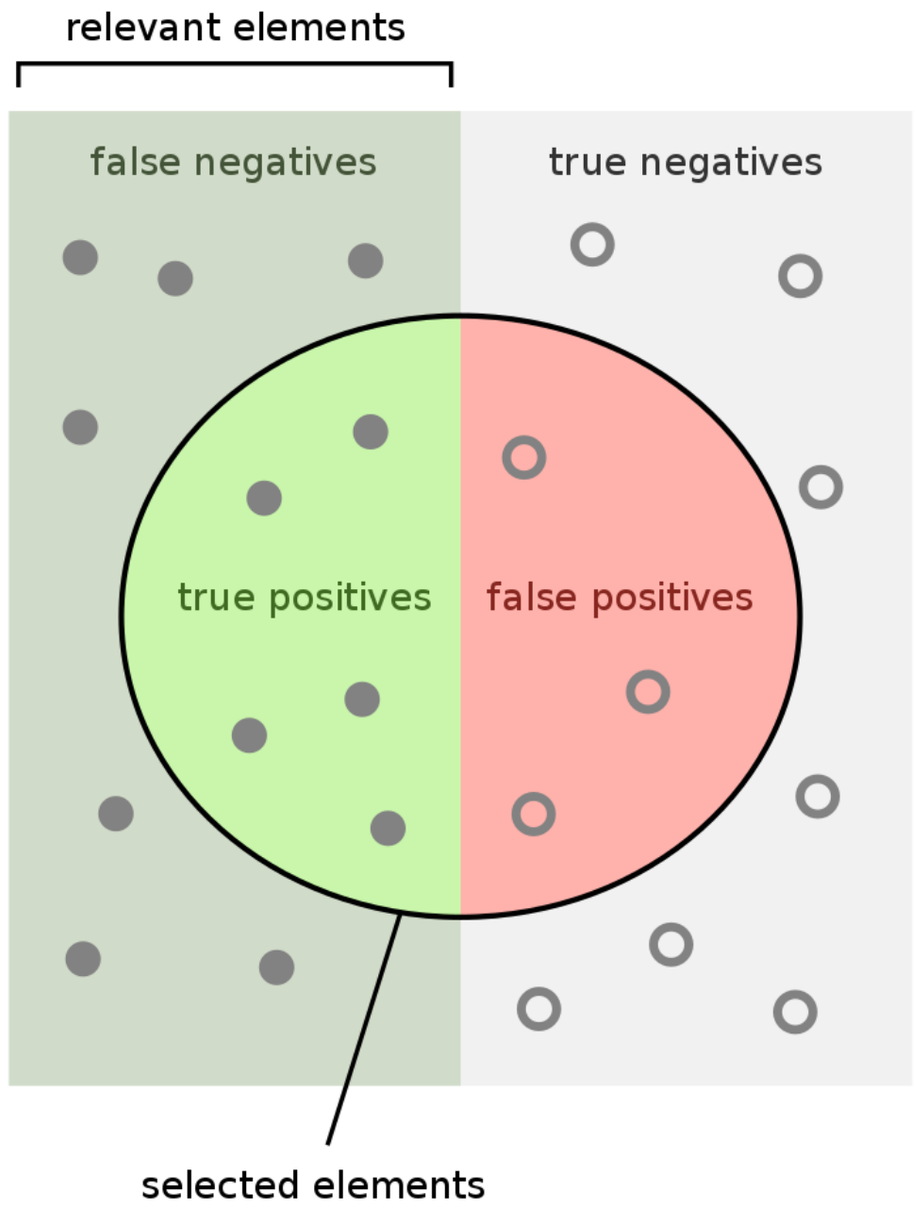
- Aside from accuracy, it is important to know how the model make wrong predictions.



Confusion Matrix

- Aside from accuracy, it is important to know how the model make wrong predictions.

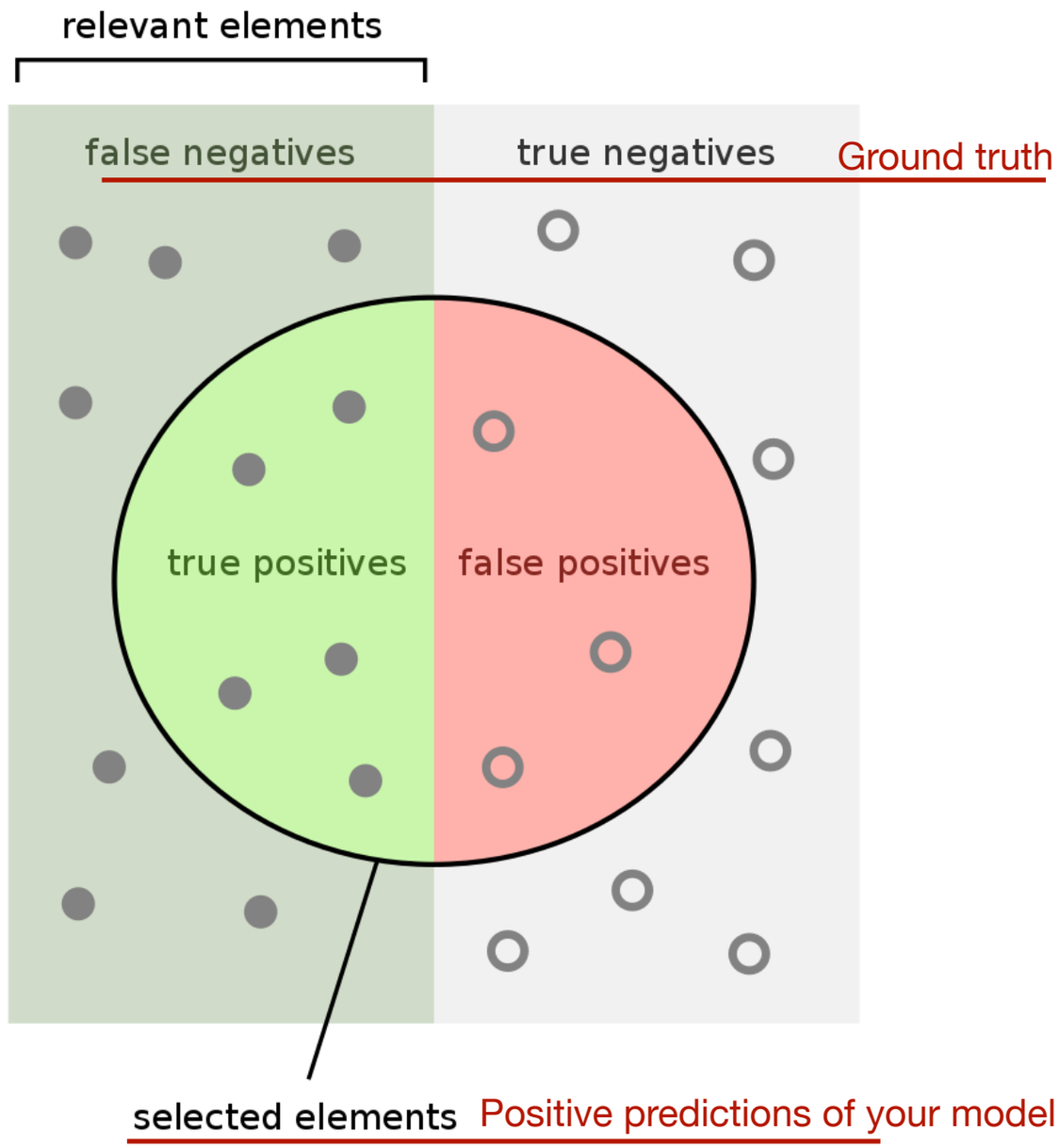
- In binary classification, confusion matrix is a common tool to analyze the predictions.



Confusion Matrix

- Aside from accuracy, it is important to know how the model make wrong predictions.

- In binary classification, confusion matrix is a common tool to analyze the predictions.



Confusion Matrix

- Aside from accuracy, it is important to know how the model make wrong predictions.

- In binary classification, confusion matrix is a common tool to analyze the predictions.

		Predicted class	
		P'	N'
Actual Class	P	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

Confusion Matrix

- Aside from accuracy, it is important to know how the model make wrong predictions.

- In binary classification, confusion matrix is a common tool to analyze the predictions.

- Many metrics are derived from the confusion matrix.

		Predicted class	
		P'	N'
Actual Class	P	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

Confusion Matrix

- Aside from accuracy, it is important to know how the model make wrong predictions.

- In binary classification, confusion matrix is a common tool to analyze the predictions.

- Many metrics are derived from the confusion matrix.

- e.g.

$$TPR = \frac{TP}{TP + FN} \qquad FPR = \frac{FP}{FP + TN}$$

		Predicted class	
		P'	N'
Actual Class	P	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

Outline

- Brief Review: Logistic Regression
  - Maximum likelihood in Logistic Regression
  - Implementation
- Common Evaluation Metrics for Binary Classification
  - Confusion Matrix
  - Soft Classifiers - ROC Curve

ROC Curve

- ROC curve analyze the performance for every threshold in soft classifiers.

1	
1	
0.87	θ
0.64	$\Downarrow$
⋮	
−0.88	
−0.93	
−1	

ROC Curve

- ROC curve analyze the performance for every threshold in soft classifiers.

- X-axis: FPR

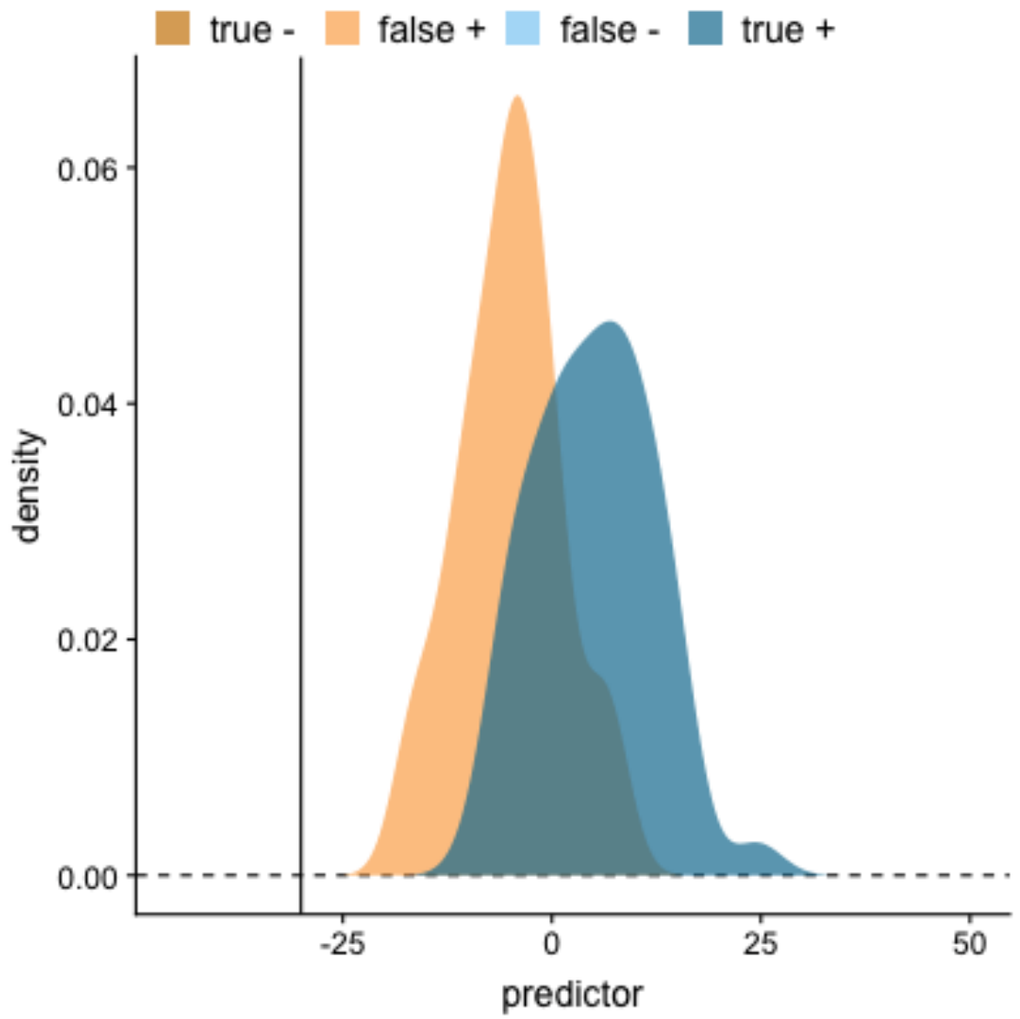
$$TPR = \frac{TP}{TP + FN}$$

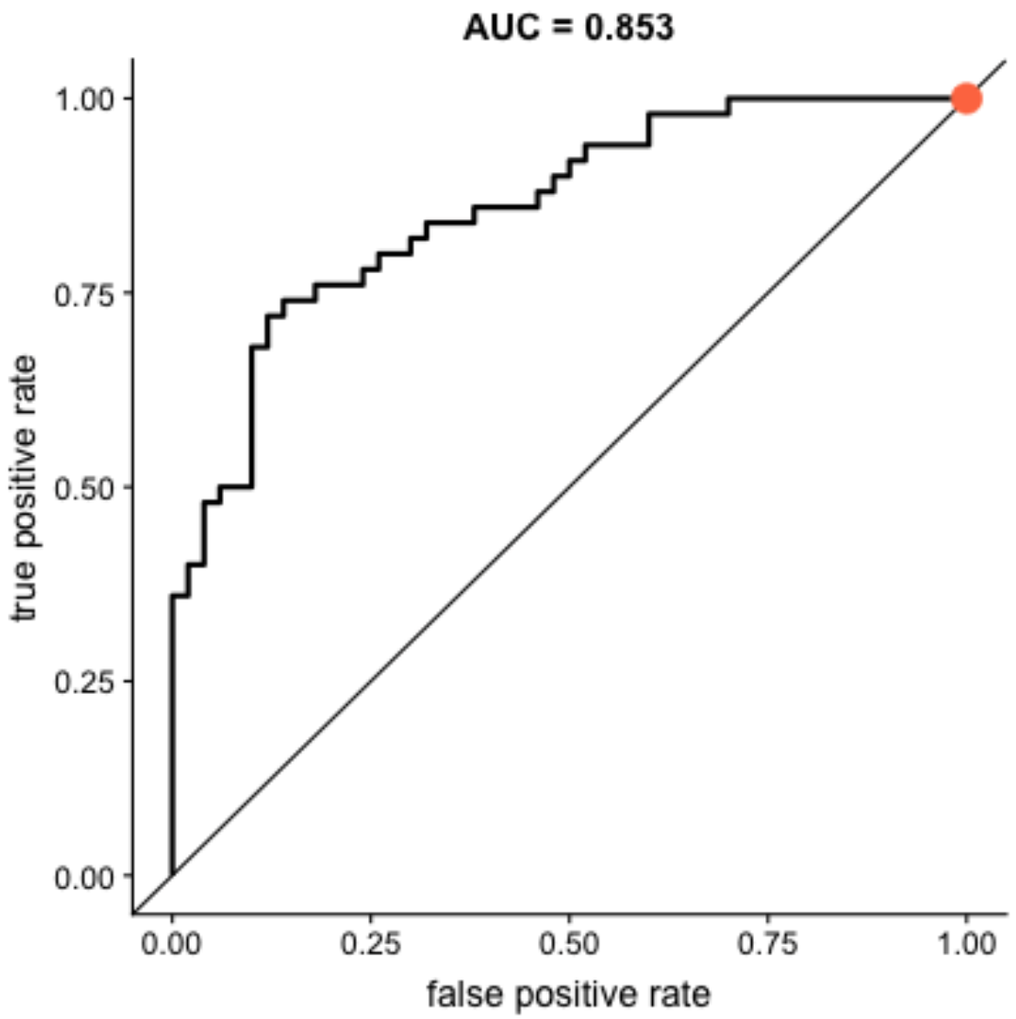
- Y-axis: TPR

$$FPR = \frac{FP}{FP + TN}$$

1	
1	
0.87	θ
0.64	$\Downarrow$
⋮	
−0.88	
−0.93	
−1	

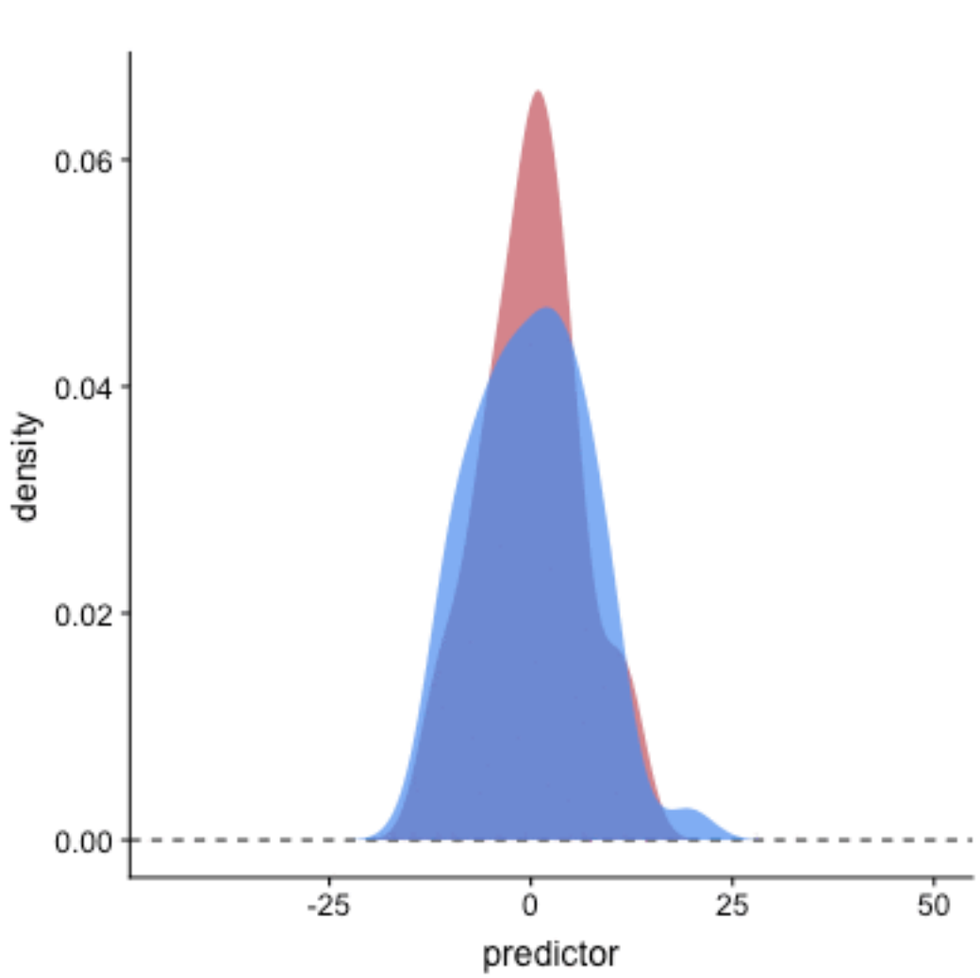
ROC Curve

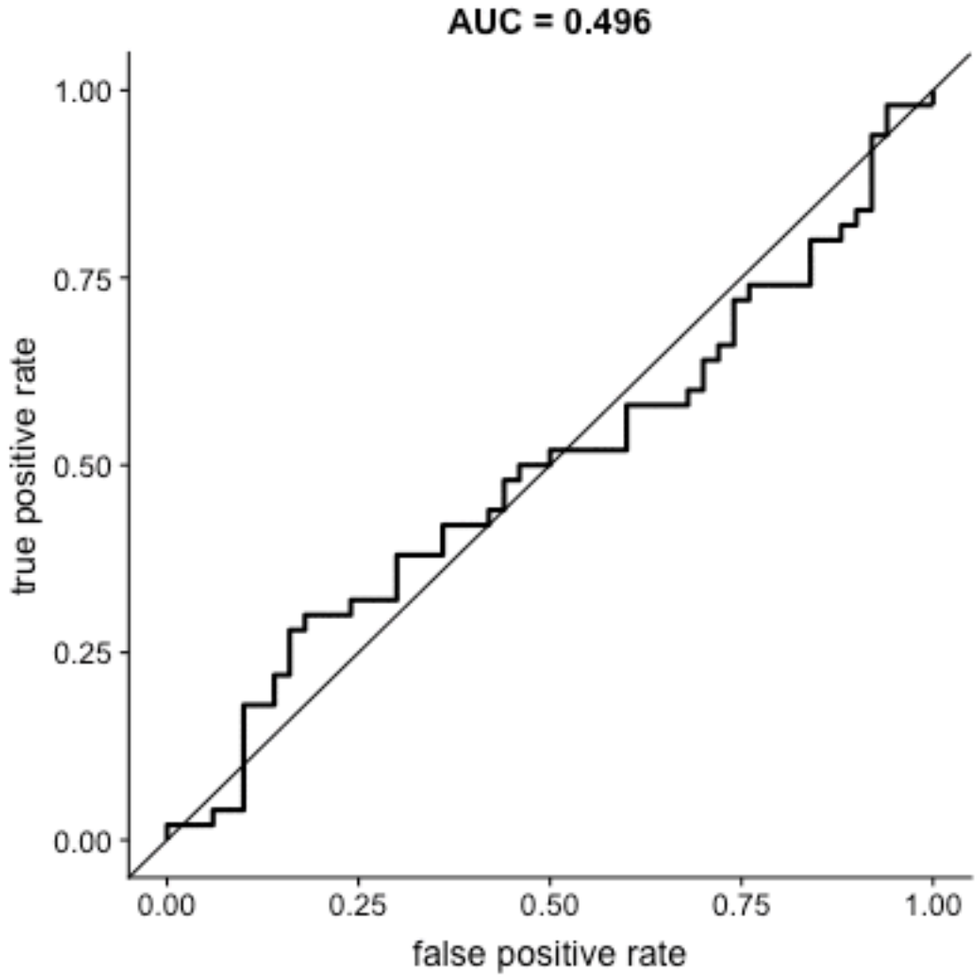

true -
false +
false -
true +
0.06
0.04
0.02
0.00
density
-25
0
25
50
predictor


AUC = 0.853
1.00
0.75
0.50
0.25
0.00
true positive rate
0.00
0.25
0.50
0.75
1.00
false positive rate

ROC Curve

- What is best ROC curve?





Homework

- Homework: Lab 6

  - Lab 6: Logistic Regression, Metrics

- Bonus: Lab 7, Lab 8

  - Lab 7: Support Vector Machine, k-Nearest Neighbors

  - Lab 8: Cross Validation, Ensemble

Homework

- Deadline: 10/20 23:59 (Tue)

  - Due to the heavy workloads, we have extended the deadline.

Reference

- https://bookdown.org/ccwang/medical_statistics6/section-43.html
- https://bookdown.org/ccwang/medical_statistics6/bernoulli.html
- https://bookdown.org/ccwang/medical_statistics6/binomial.html
- https://bookdown.org/ccwang/medical_statistics6/likelihood-definition.html
- https://en.wikipedia.org/wiki/Sensitivity_and_specificity
- https://github.com/dariyasydykova/open_projects/tree/master/ROC_animation