Lab6: Logistic Regression and Metrics

Hao-Lun Sun & DataLab

2021.10.14

Outline

- Brief Review: Logistic Regression
  - Maximum likelihood in Logistic Regression
  - Implement

- Common Evaluation Metrics for Binary Classification
  - Confusion Matrix
  - Soft Classifiers – ROC Curve

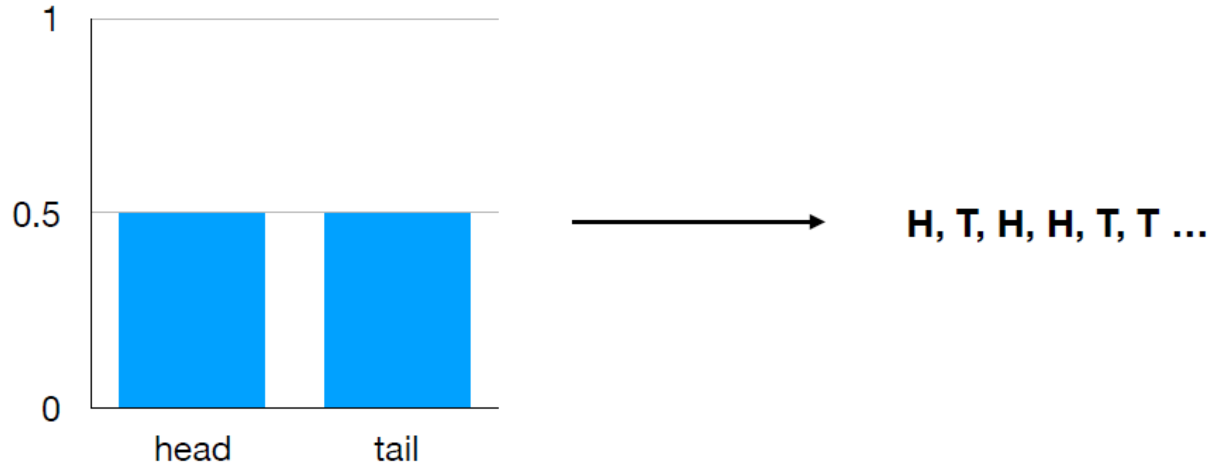
Outline

- Brief Review: Logistic Regression
  - Maximum likelihood in Logistic Regression
  - Implement

- Common Evaluation Metrics for Binary Classification
  - Confusion Matrix
  - Soft Classifiers – ROC Curve

Maximum Likelihood

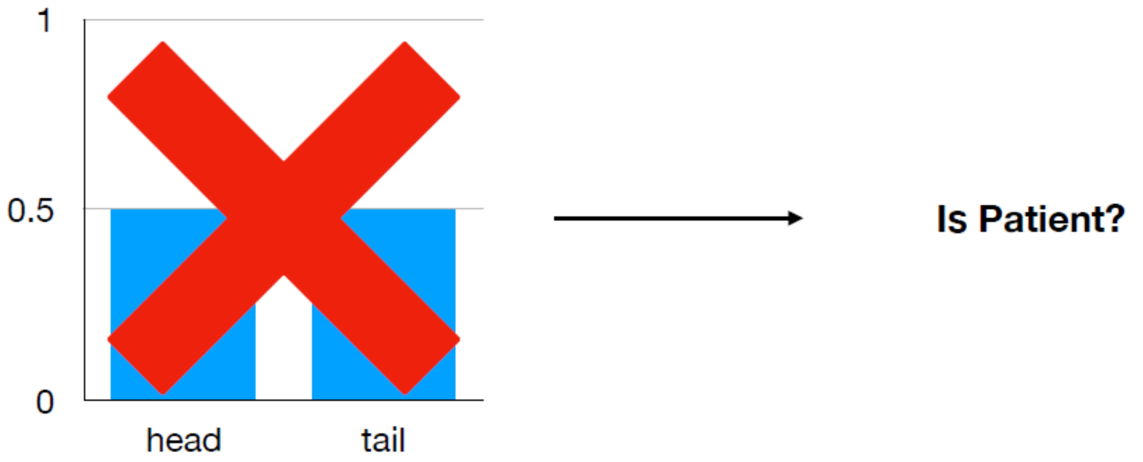
- Flipping coin: We have already known ground truth distribution. For example, $P(x = head) = \frac{1}{2}$ and $P(x = tail) = \frac{1}{2}$.



Maximum Likelihood

- Flipping coin: We have already known ground truth distribution. For example, $P(x = head) = \frac{1}{2}$ and $P(x = tail) = \frac{1}{2}$.

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



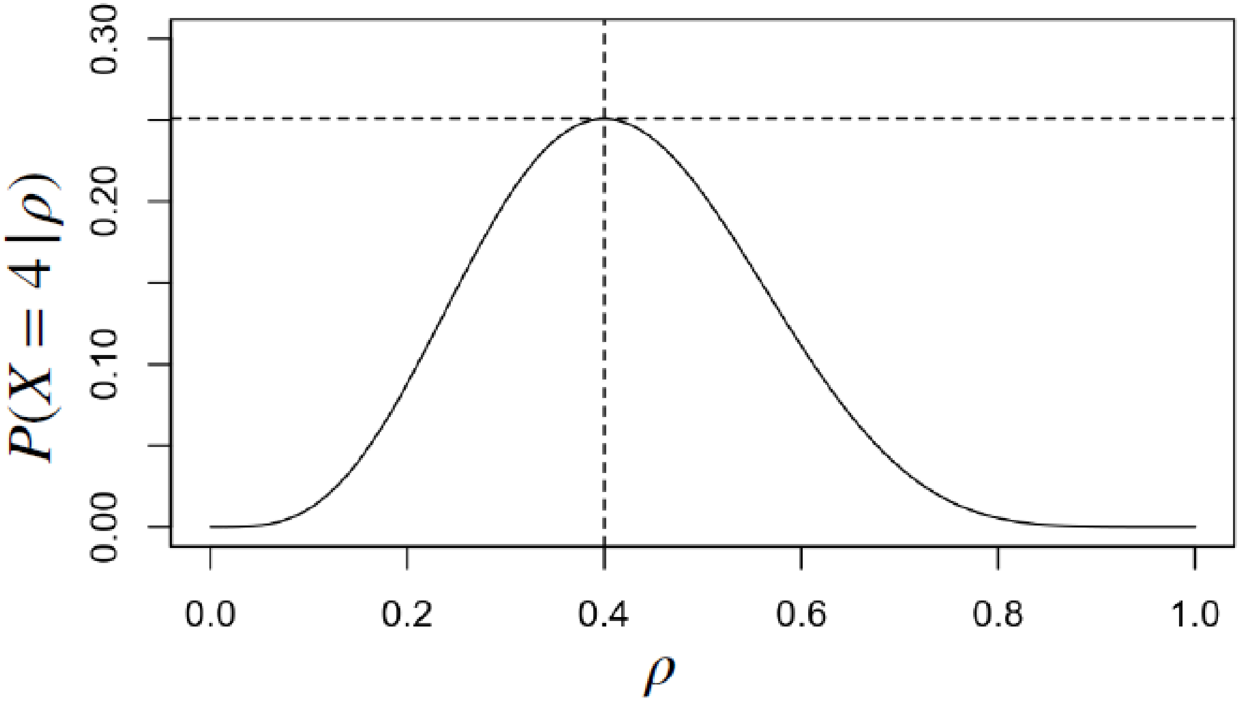
Maximum Likelihood

- The process to approximate the distribution:
  - First, we assume the proportion of people diagnosed with a disease follow Binomial distribution, e.g., $X \sim Bin(A, \rho)$.
  - A is the number of person that diagnosed, ρ is illness rate.
  - If there are 4 patients out of 10 people, the number of Binomial 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

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



Outline

- Brief Review: Logistic Regression
  - Maximum likelihood in Logistic Regression
  - Implement

- Common Evaluation Metrics for Binary Classification
  - Confusion Matrix
  - Soft Classifiers – ROC Curve

Logistic Regression

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

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

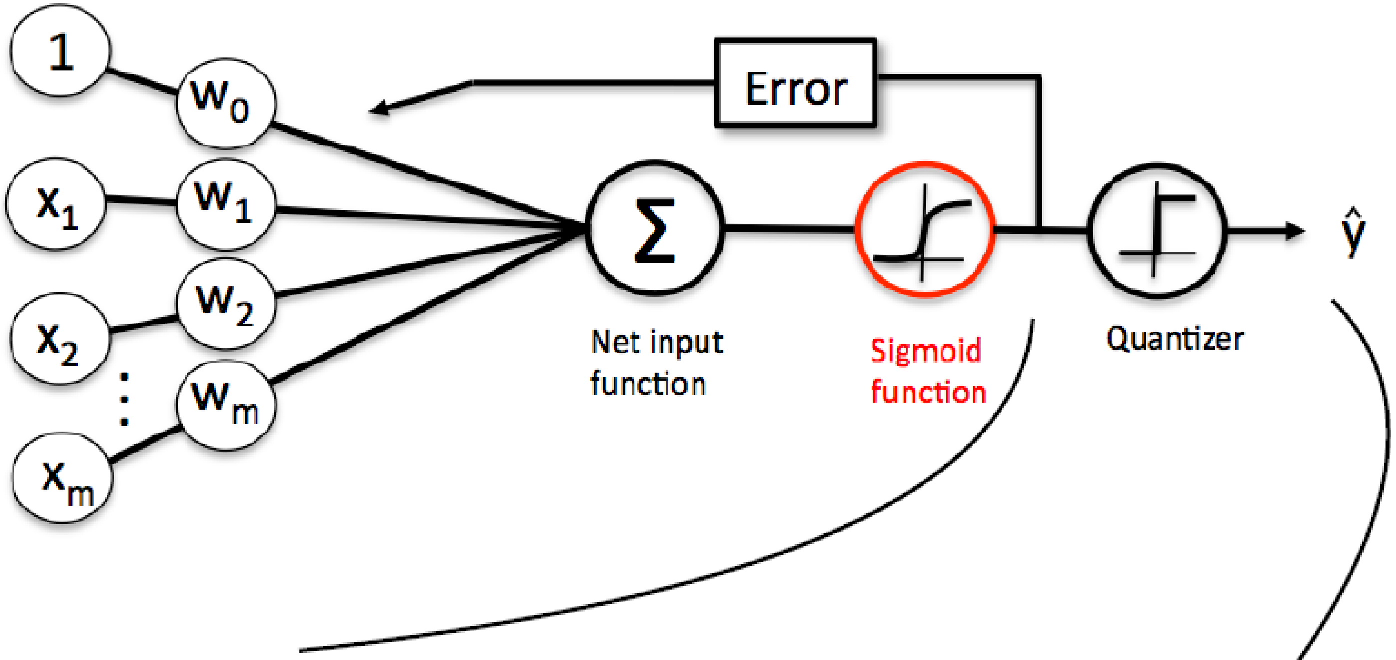
- Update with gradient decent:

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

- 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



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

Soft prediction

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

Label prediction

Outline

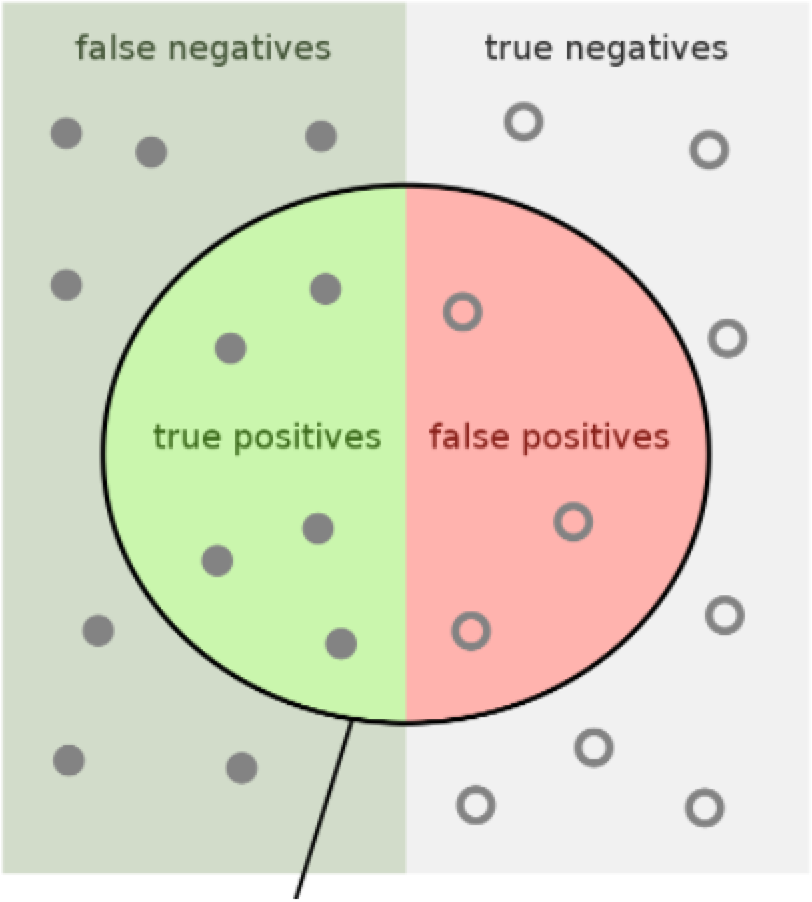
- Brief Review: Logistic Regression
  - Maximum likelihood in Logistic Regression
  - Implement

- Common Evaluation Metrics for Binary Classification
  - Confusion Matrix
  - Soft Classifiers – ROC Curve

Confusion Matrix

- It is important to know how the model make wrong prediction.

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



Confusion Matrix

- It is important to know how the model make wrong prediction.

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

- Other metrics we can use:

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

		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
  - Implement

- 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.

- In X-axis: FPR

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

- In Y-axis: TPR

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

1	
1	
0.87	θ
0.64	$\Downarrow$
$\vdots$	
-0.88	
-0.93	
-1	

ROC Curve

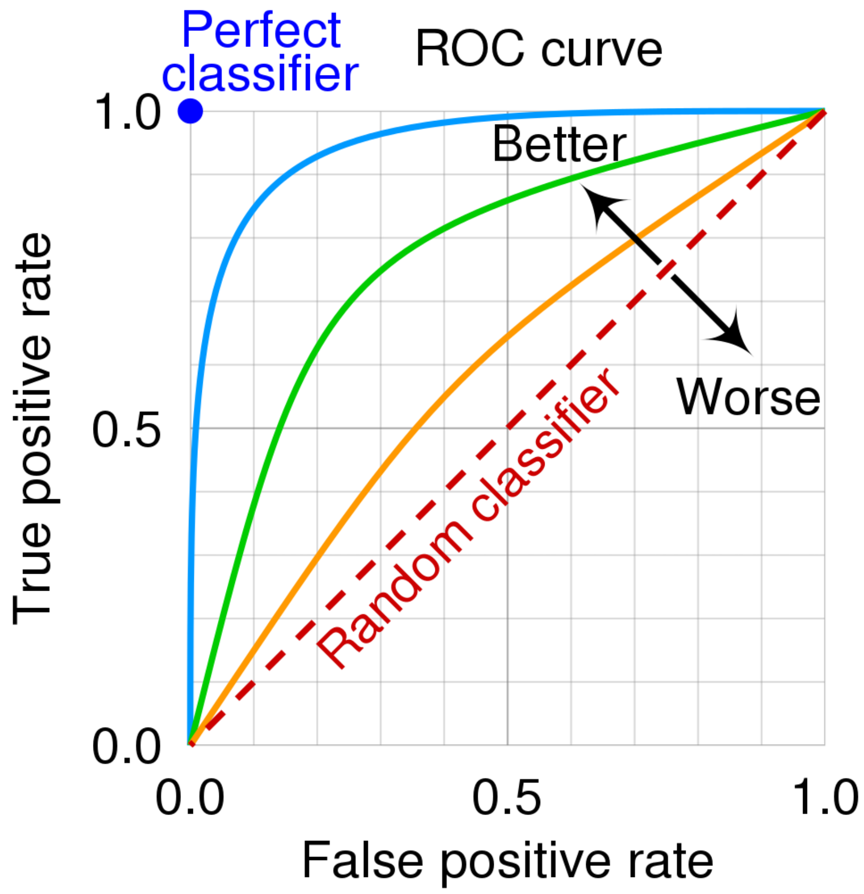
- ROC curve analyze the performance for every threshold in soft classifiers.

- In X-axis: FPR

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

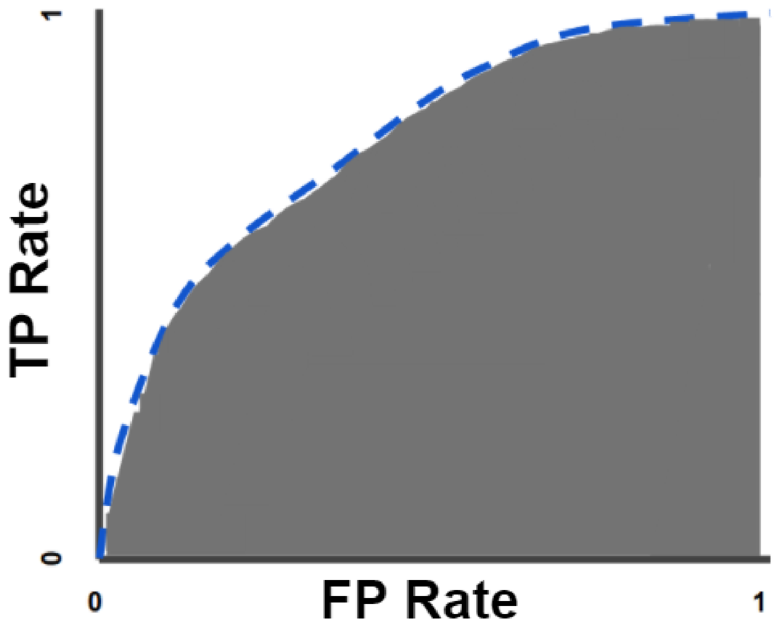
- In Y-axis: TPR

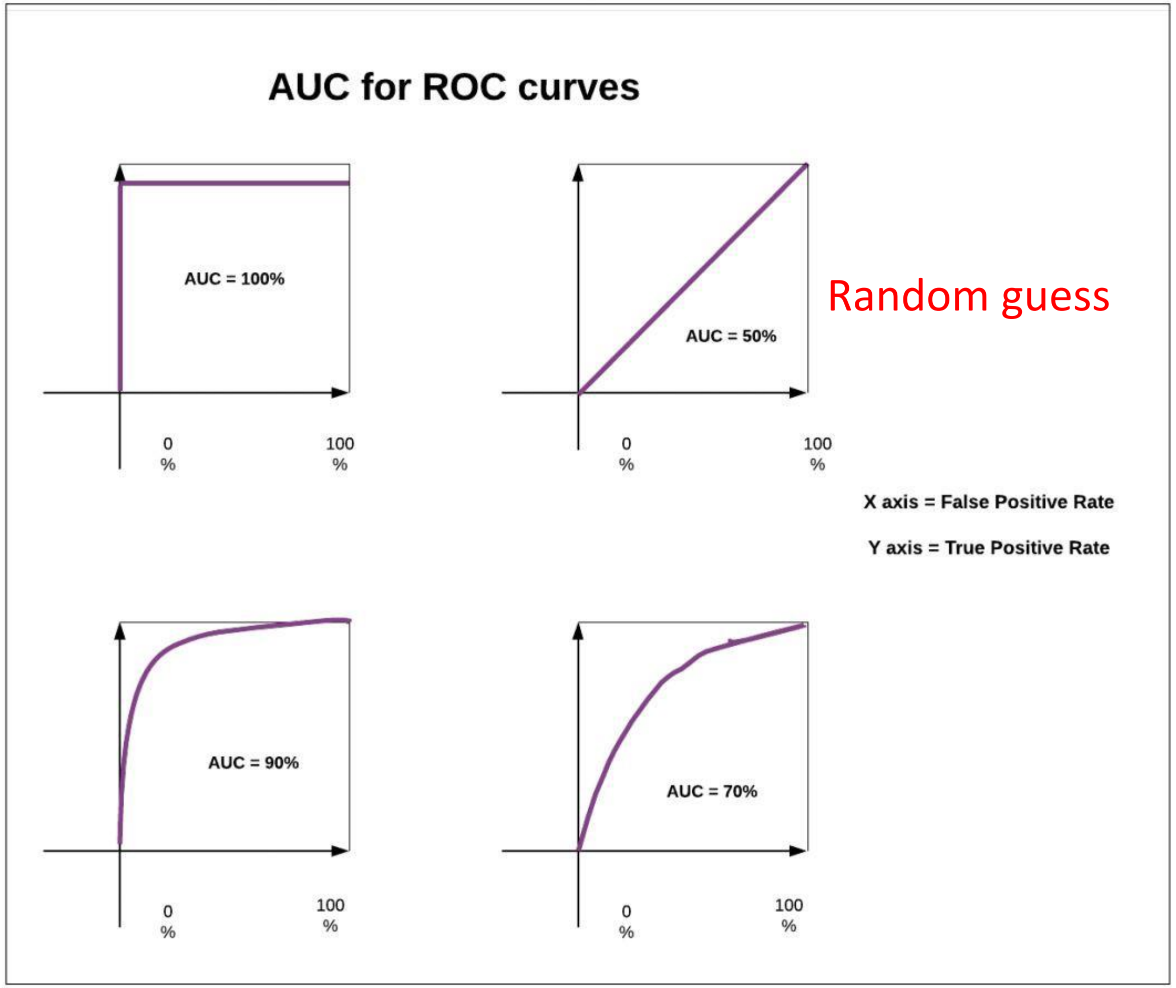
$$TPR = \frac{TP}{TP + FN}$$



ROC Curve

- AUC - Area Under the ROC Curve.
  - ROC can be quantified using AUC.





https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc
https://medium.com/acing-ai/what-is-auc-446a71810df9

Homework

- Homework: Lab06
  - Lab06: Logistic Regression, Metrics

- Bonus: Lab07 && Lab08
  - Lab07: Support Vector Machine, k-Nearest Neighbors
  - Lab08: Cross Validation, Ensemble

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://commons.wikimedia.org/w/index.php?curid=109730045
- https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc
- https://medium.com/acing-ai/what-is-auc-446a71810df9
- https://github.com/dariyasydykova/open_projects/tree/master/ROC_animation