Week 7: Model Training and Evaluation

Ekarat Rattagan

September 12, 2025

Which model is the best choice?

$$h_\theta(x) = \theta_0 + \theta_1 x_1$$

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2^2$$

...

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2^2 + ... + \theta_9 x_9^9$$

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2^2 + ... + \theta_9 x_9^9 + \theta_{10} x_{10}^{10}$$

Two Methods in Training process (1/2)

- **Hold-out Method**
  - 80% Train set
  - 20% Test set
- Pros:
  - Simple
  - Fast
- Cons:
  - Overfitting

Two Methods in Training process (2/2)

- **Cross validation Method**
  - Training set: 60%
  - Validation set: 20%
  - Test set: 20%
- Pros:
  - Variance (Overfitting) is reduced
- Cons:
  - Slow

Which one reduces the error caused by overfit or underfit?

- Get more training data
- Try smaller sets of features
- Try getting additional features
- Try adding polynomial features
- Try decreasing λ (Ridge penalty)
- Try increasing λ (Ridge penalty)

Bias vs Variance

- **Underfit:** High Bias (Linear Model: $\theta_0 + \theta_1 x$)
- **Good Fit:** Proper Balance (Quadratic Model: $\theta_0 + \theta_1 x + \theta_2 x^2$)
- **Overfit:** High Variance (Higher-order Polynomial Model:

  $\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$)

Diagnosis: Bias vs Variance (Polynomial Degree)

- Low-degree polynomial: Underfitting
- Moderate-degree polynomial: Good Fit
- High-degree polynomial: Overfitting

Diagnosis: Bias vs Variance (Regularization Parameter)

- Ridge function:

$$J(\theta) = \frac{1}{N}\sum_{i=1}^{N}(h_\theta(x_i) - y_i)^2 + \lambda\sum_{j=1}^{d}\theta_j^2 \tag{1}$$

- Increasing λ reduces variance but increases bias
- Decreasing λ reduces bias but increases variance

Effect of Training Data Size

- Increasing training data reduces variance but not bias

Error Metrics

- Confusion Matrix
- Precision, Recall, Accuracy
- F1-Score
- ROC, AUC

Confusion Matrix

	Actual Positive	**Actual Negative**
Predicted Positive	TP (True Positive)	FP (False Positive)
Predicted Negative	FN (False Negative)	TN (True Negative)

Key Metrics:

- Accuracy: $\frac{TP+TN}{TP+FP+TN+FN}$
- Recall: $\frac{TP}{TP+FN}$
- Precision: $\frac{TP}{TP+FP}$
- Sensitivity (TPR): $\frac{TP}{TP+FN}$
- Specificity (TNR): $\frac{TN}{TN+FP}$
- False Positive Rate (FPR): $\frac{FP}{FP+TN}$
- Matthews Correlation Coefficient (MCC):

$$\frac{TP\times TN-FP\times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

Accuracy, Recall, Precision

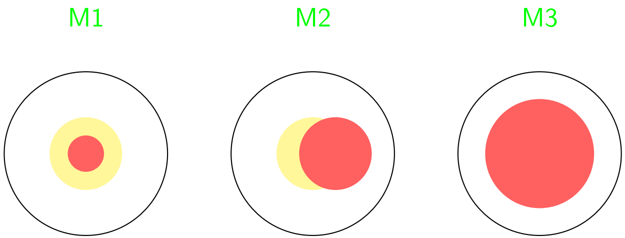
id	y	$\hat{y}_{\text{prob}}$	$\hat{y}_{\text{threshold} \geq 0.5}$	Accuracy	Recall	Precision
1	0	0.5	1			
2	1	0.9	1			
3	0	0.7	1			
4	1	0.7	1			
5	1	0.3	0			
6	0	0.4	0			
7	1	0.5	1			

Accuracy, Recall, Precision

id	y	$\hat{y}_1$	$\hat{y}_2$	$\hat{y}_3$	Accuracy	Recall	Precision
1	0	0	0	1			
2	0	0	0	1			
3	0	0	0	1			
4	0	0	0	1			
5	0	0	0	1			
6	0	0	0	1			
7	0	0	0	1			
8	0	0	0	1			
9	1	0	1	1			
10	1	0	0	1			

Tumor detection (Accuracy, Recall, Precision)

- Tumor area (Positive)
- Predicted area (Negative)



F1 Score

How to average precision and recall value?

$$\text{Average (Arithmetic Mean)} = \frac{P + R}{2}$$

$$F_1 \text{ score (Harmonic Mean)} = \frac{2 \times P \times R}{P + R}$$

Model	Precision (P)	Recall (R)	Avg $\frac{P+R}{2}$	F_1
1	0.5	0.4		
2	0.7	0.1		
3	0.02	1.0		

Receiver operating characteristic (ROC)

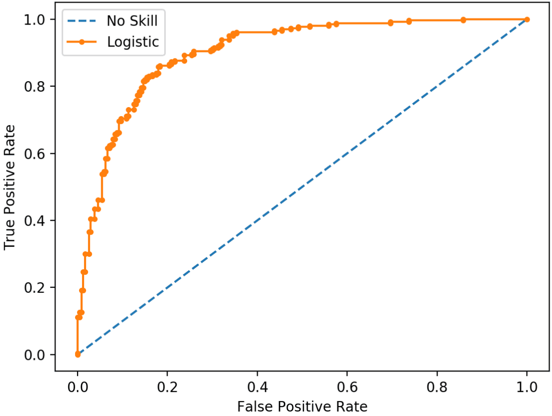


Figure: Receiver operating characteristic (ROC)

Ref: `https://machinelearningmastery.com`

Precision-Recall Curve

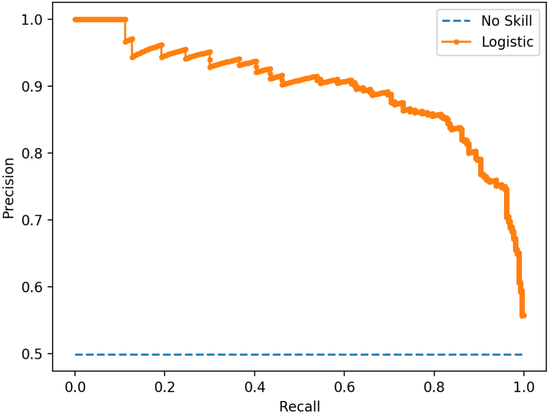


Figure: Precision-Recall Curve (PR-curve)

Ref: `https://machinelearningmastery.com`

Reference

- Raschka,S.(2018).Modelevaluation,modelselection, andalgorithmselectioninmachinelearning.arXivpreprint arXiv:1811.12808.