EE353: Elastic Net and Binary Classification Metrics

Sravan K Suresh

September 19, 2024

1. Elastic Net Regularization

Elastic Net combines both **L1 (Lasso)** and **L2 (Ridge)** regularization:

$$\min_{\beta} \left\{ \frac{1}{2n} \|y - X\beta\|_2^2 + \lambda \left(\alpha \|\beta\|_1 + (1-\alpha)\|\beta\|_2^2 \right) \right\}$$

- $\alpha \in [0, 1]$: mixing parameter. $\alpha = 1$ gives Lasso, $\alpha = 0$ gives Ridge.
- λ: regularization strength.

Elastic Net is useful when there are highly correlated features, as it encourages a grouping effect.

2. Binary Classification Metrics

Confusion Matrix

The confusion matrix provides insights into the performance of a classification model:

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

Accuracy

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- Proportion of correct predictions.

Sensitivity (Recall) or True Positive Rate (TPR)

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

- Measures the ability to correctly predict positive instances.

Specificity or True Negative Rate (TNR)

$$\text{Specificity} = \frac{TN}{TN + FP}$$

- Measures the ability to correctly predict negative instances.

Precision

$$\text{Precision} = \frac{TP}{TP + FP}$$

- Proportion of positive predictions that are actually positive.

F1 Score

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

- Harmonic mean of precision and recall. Useful when dealing with imbalanced data.

AUC-ROC (Area Under the Receiver Operating Characteristic Curve)

- **ROC Curve**: Plots TPR (Sensitivity) against FPR (1 - Specificity) at various threshold levels.
- **AUC (Area Under Curve)**: Measures the area under the ROC curve. A perfect model has an AUC of 1, while a random model has an AUC of 0.5.

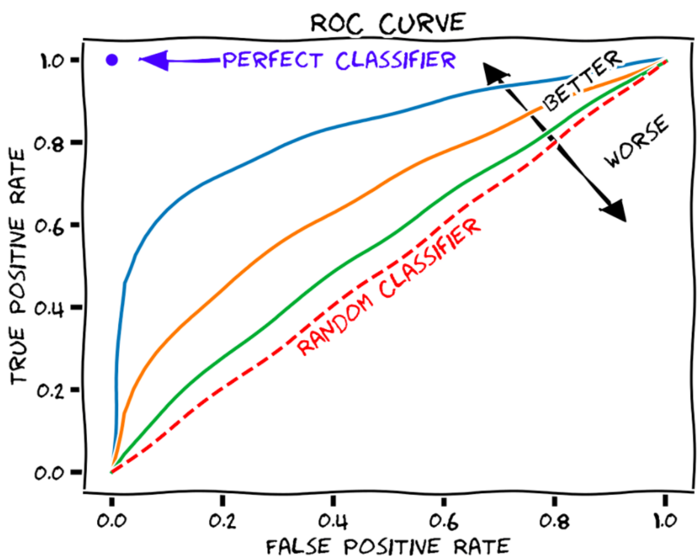

Figure 1: ROC Curves

3. Asymmetric Risk

In some scenarios, the cost of different types of errors (FP, FN) is unequal:
Example: In medical diagnosis, a false negative (missed diagnosis) may be riskier than a false positive (unnecessary further testing).

Asymmetric risk is addressed by adjusting classification thresholds or using different loss functions that penalize errors unequally:

$$\text{Risk} = \text{Cost(FP)} \cdot P(FP) + \text{Cost(FN)} \cdot P(FN)$$

NOTE: Classifier tuning can minimize the risk rather than maximizing accuracy.