Evaluation Metrics for Classification - Part 2

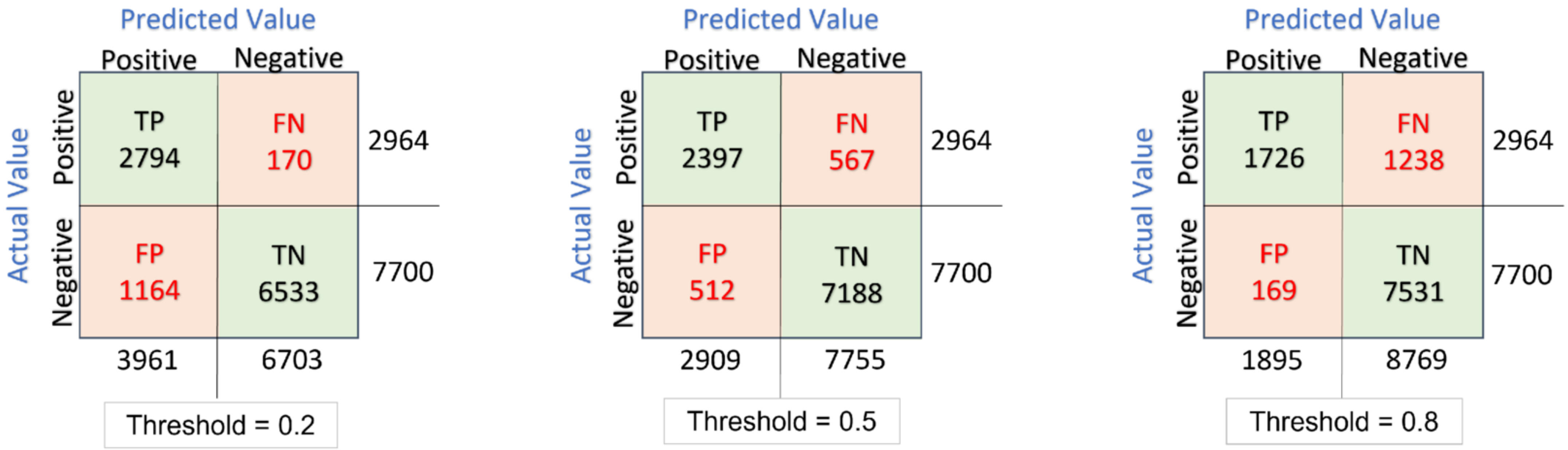
Logistic Regression II

Learning Objectives

1. Understand the Area Under the Curve (AUC) metric used to evaluate classification models.
2. Apply the evaluation metrics to compare models and check for overfitting.
3. Apply the Logistic Regression model on sample data and make data-driven recommendations.

Sensitivity/Specificity Tradeoff

- Evaluation metrics based on confusion matrix use a single cut-off threshold, e.g. 0.5. Varying the threshold, changes the counts in the confusion matrix.



- At a low threshold of 0.2, model predicts higher no. of Positives. False Positives increase and False Negatives decrease.
- At a high threshold of 0.8, model predicts lower no. of Positives. False Positives decrease and False Negatives increase.

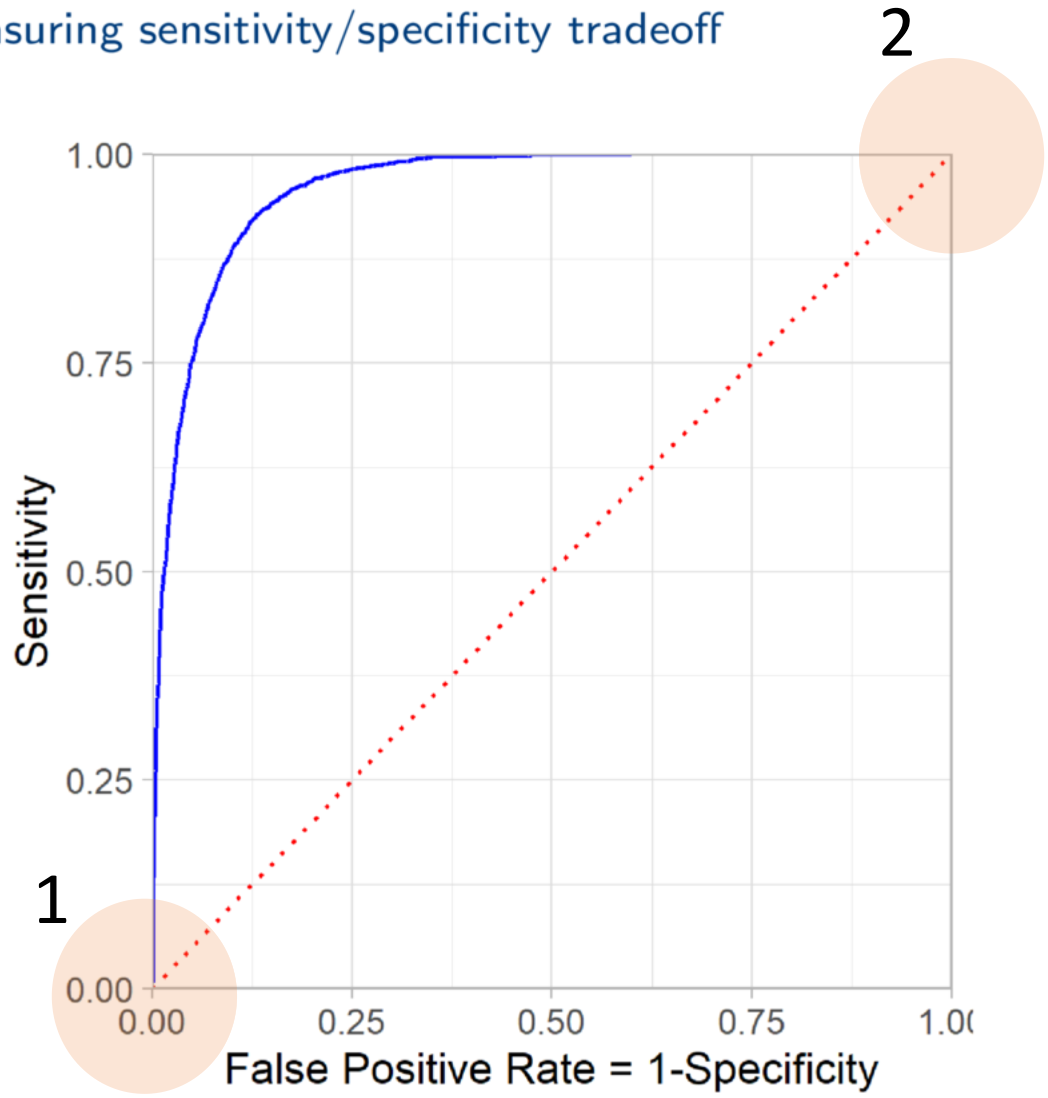
Sensitivity/Specificity Tradeoff

- Sensitivity: When actual value is Yes, how often does the model predict Yes? Ideally, Sensitivity $= 1$.
- Specificity: When actual value is No, how often does the model predict No? Ideally Specificity $= 1$, or the false positive rate, $1 -$ Specificity $= 0$.
- When cut off threshold is lowered, sensitivity increases, but specificity reduces, and therefore, false positive rate increases, and vice-versa for higher threshold.

	Threshold = 0.2	Threshold = 0.5	Threshold = 0.8
Sensitivity	0.94	0.81	0.58
1 - Specificity	0.15	0.07	0.02

ROC Curve

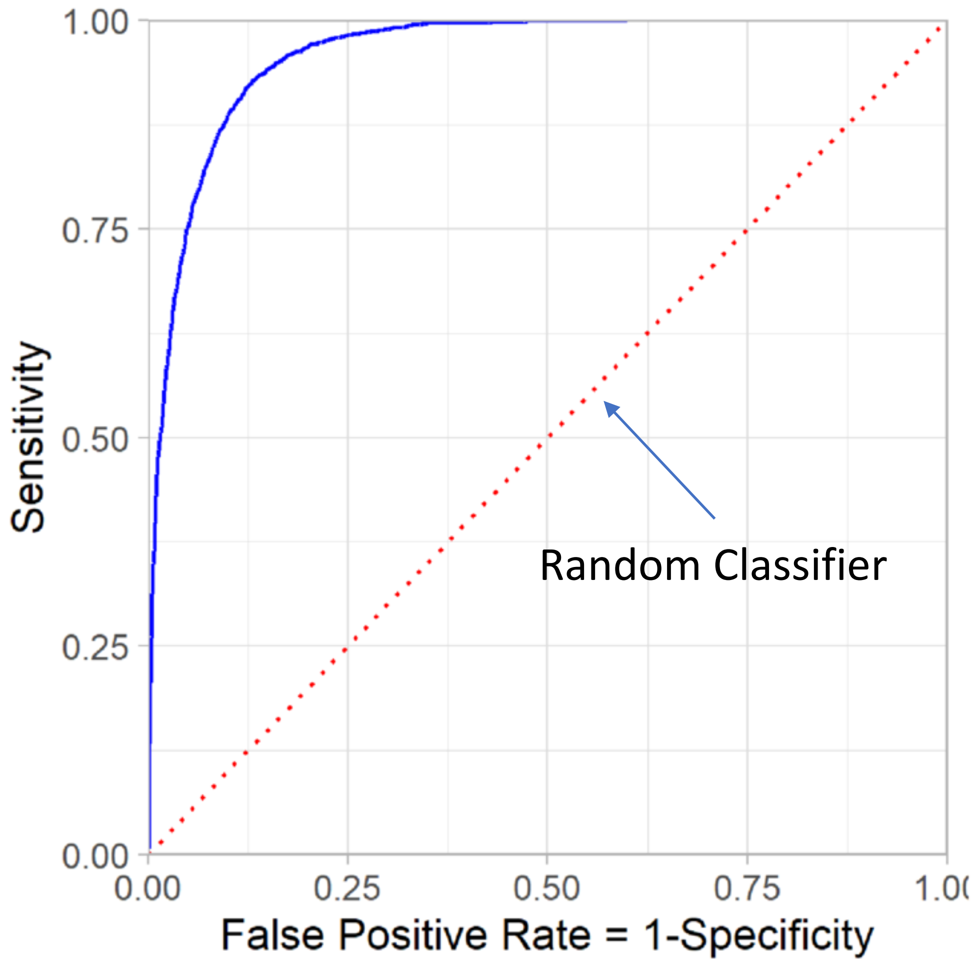
Measuring sensitivity/specificity tradeoff



- ROC Curve captures the tradeoff between sensitivity and specificity of the Logistic Regression model.
  1. (0,0) on ROC: When threshold = 1, model classifies all data as Negatives, and does not predict any Positives. So, sensitivity = 0, and false positive rate = 0.
  2. (1,1) on ROC: When threshold = 0, model classifies all data as Positives, and does not predict any Negatives. So, sensitivity = 1, and false positive rate = 1.

ROC Curve

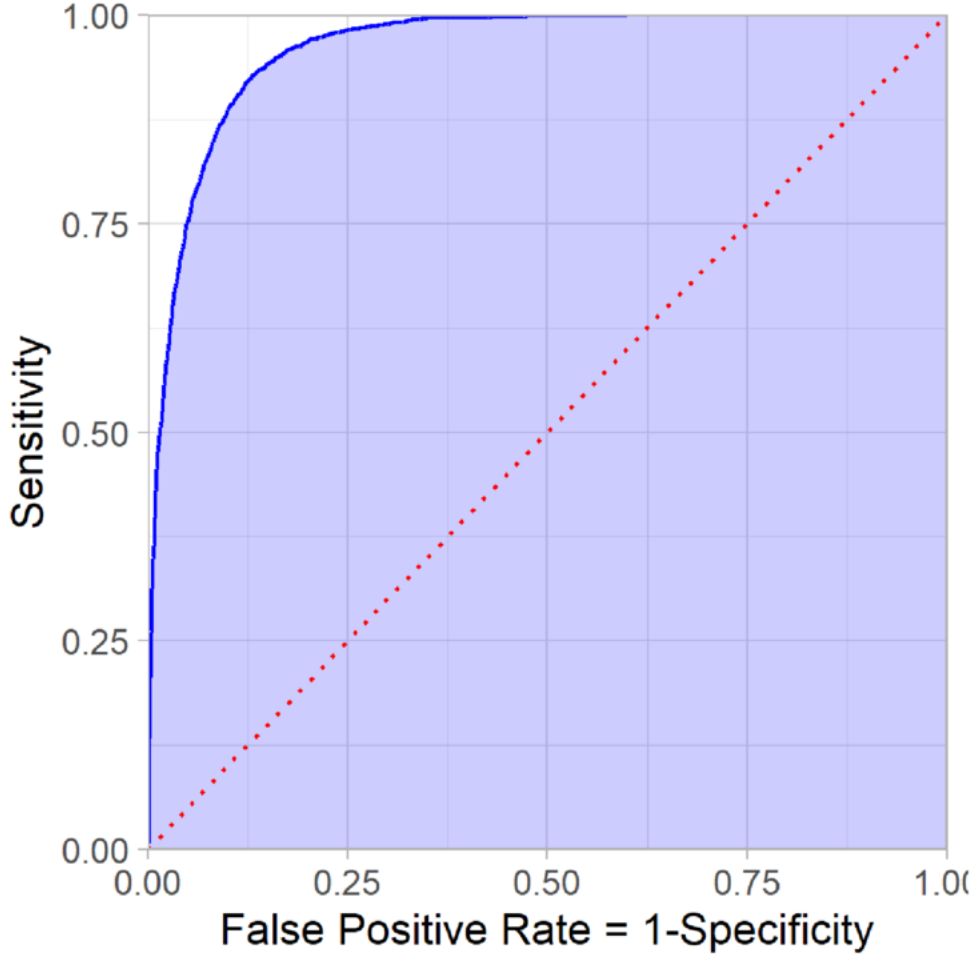
Measuring sensitivity/specificity tradeoff



- Diagonal dotted line represents a classification model that is no better than random chance.
- Extremely effective classification model hugs the upper left corner of ROC curve - correcly identifies lots of Positives without misclassifying lots of Negatives as Positives.

Area Under the Curve (AUC) of ROC Curve

Measuring sensitivity/specificity tradeoff



- AUC represents the area underneath the ROC curve. Here, AUC $= 0.96$.
- For a perfect classifier, AUC $= 1$.
- For a completely ineffective classifier based on random chance (diagonal line), AUC $= 0.5$.

AUC of ROC Curve

Code

- To generate the ROC curve, we compute all pairs of sensitivity and specificity values for the entire range of predicted probabilities.

```
idx <- order(-pp$y_prob)
sensitivity <- cumsum(pp$default[idx] == "Yes")/sum(pp$default == "
    Yes")
specificity <- (sum(pp$default == "No") - cumsum(pp$default[idx] == "
    No"))/sum(pp$default == "No")
roc_df <- data.frame(sensitivity = sensitivity, specificity =
    specificity)
```

AUC of ROC Curve

Code

```
ggplot(roc_df, aes(x=1-specificity, y=sensitivity)) +
   geom_line(color = "blue") +
   geom_ribbon(aes(xmin=0, ymin=0, xmax=1, ymax=sensitivity),
               fill="blue", alpha = 0.2) +
   scale_x_continuous(expand=c(0,0)) +
   scale_y_continuous(expand = c(0,0)) +
   geom_line(data = data.frame(x=(0:100)/100),
             aes(x=x, y=x), linetype = "dotted", color = "red") +
   labs(x="False Positive Rate = 1-Specificity", y="Sensitivity") +
   theme_light()
auc <- sum(roc_df$sensitivity[-1]*diff(1-roc_df$specificity))
auc
```

```
[1] 0.9606026
```

Model Comparison

- Rule of thumb: at least 4 out of 6 evaluation metrics are above 0.8.

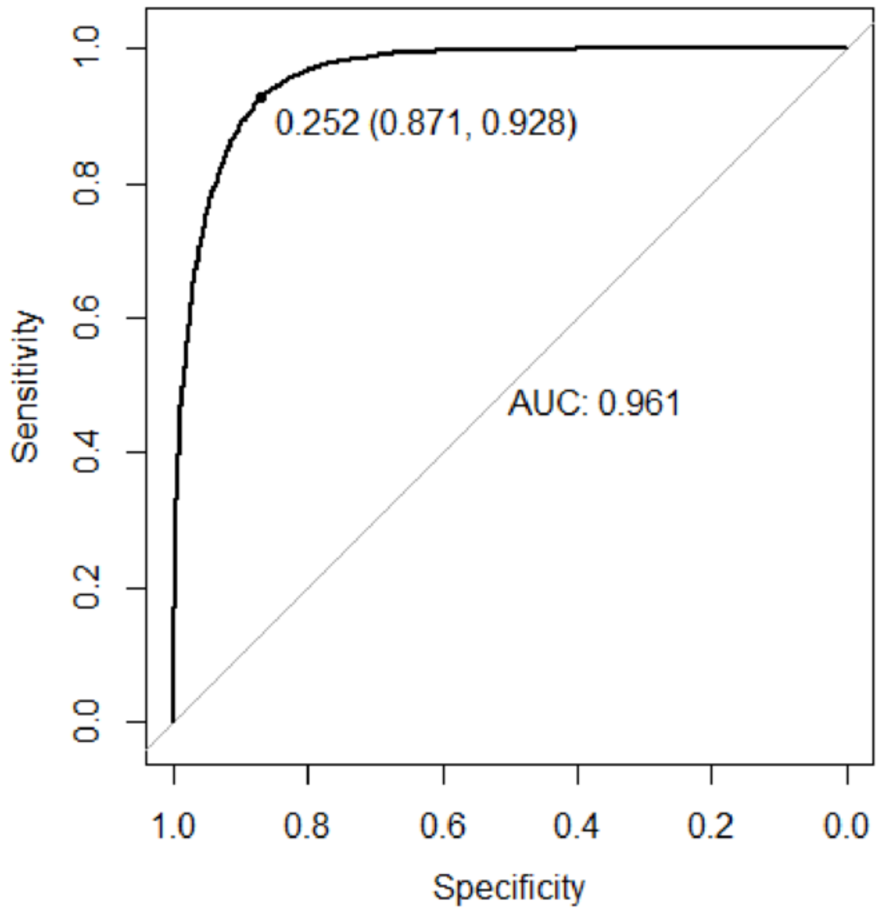
Formula	AIC	Accuracy	Sensitivity	Specificity	Precision	F1 Score	AUC
balance + income + student	4863.81	0.9	0.81	0.93	0.82	0.82	0.96
balance + student	4863.04	0.9	0.81	0.93	0.82	0.82	0.96
income + balance	4888.15	0.9	0.81	0.93	0.83	0.82	0.96
balance	4948.99	0.9	0.82	0.93	0.83	0.82	0.96

Check for Overfitting

- Apply Logistic Regression model on unseen test data and compute evaluation metrics.
- No overfitting as evaluation metrics are similar between train and test data.

Metrics	Train	Test
Accuracy	0.899	0.900
Sensitivity	0.809	0.820
Specificity	0.934	0.929
Precision	0.824	0.804
F1 Score	0.816	0.812
AUC	0.961	0.961

Optimal Cut-off Threshold



- Optimal threshold for classification can be selected using the ROC curve.
- At threshold $= 0.252$, we get the best tradeoff between Sensitivity $= 0.928$ and Specificity $= 0.871$.

Optimal Cut-off Threshold

Code

```
library(pROC)
y_true <- train$default
y_probs <- predict(model2, newdata = train, type = "response")
plot.roc(y_true, y_probs, print.auc = TRUE,
         thresholds="best", print.thres="best")
```

APPLY: Credit default data

- Suppose the bank wants predict if these two customers are likely to default:
  - Customer 1: Student with a credit card balance of $1400.
  - Customer 2: Non-student with a credit card balance pf $1350.

```
sample <- data.frame(balance=c(1400, 1350),
                     student=c("Yes", "No"))
predict(model2, newdata = sample, type = "response")
```

```
        1         2
0.2425367 0.3235812
```

- Customer 1 has a probability of default of 0.24. Customer 2 has a probability of default of 0.32.

References I

Bruce, P., Bruce, A. G., and Gedeck, P. (2020).
Practical statistics for data scientists: 50 essential concepts using R and Python.
OReilly.