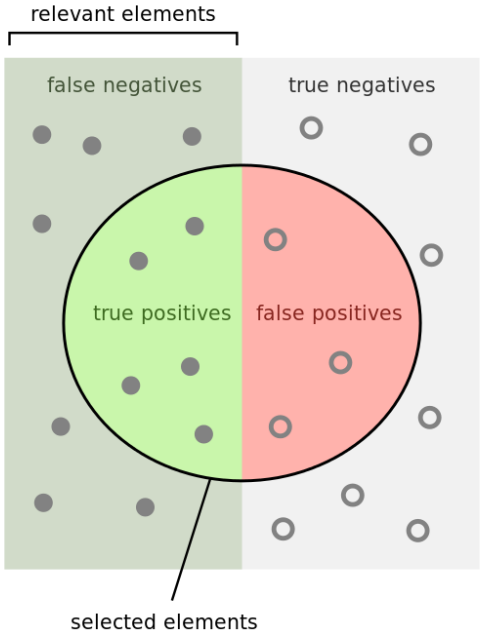
livelessons
video instruction from technology experts

# Lesson 5: Advanced Applications

## 5.6 Leveraging scikit-learn to Evaluate MLlib Models

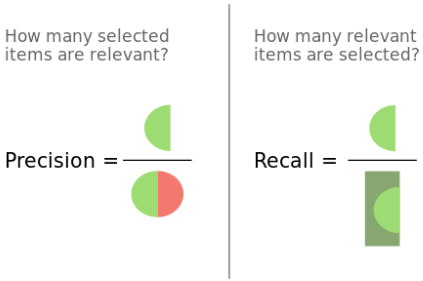
# Precision and Recall




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

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

$$F1 = \frac{2}{\frac{1}{\text{precision}} + \frac{1}{\text{recall}}}$$




# Confusion Matrix vs. Point Metrics

- Confusion matrix has fine-grained information about misclassifications
- Precision/Recall/F1 can be used in automated comparison (grid search)

# Logistic Regression: Balanced Classes

```
print(classification_report(y_true, y_pred))
```

```
             precision    recall  f1-score   support

        0.0       0.74      0.92      0.82      1375
        1.0       0.70      0.37      0.49       705

avg / total       0.73      0.73      0.71      2080
```

# Unified Platform