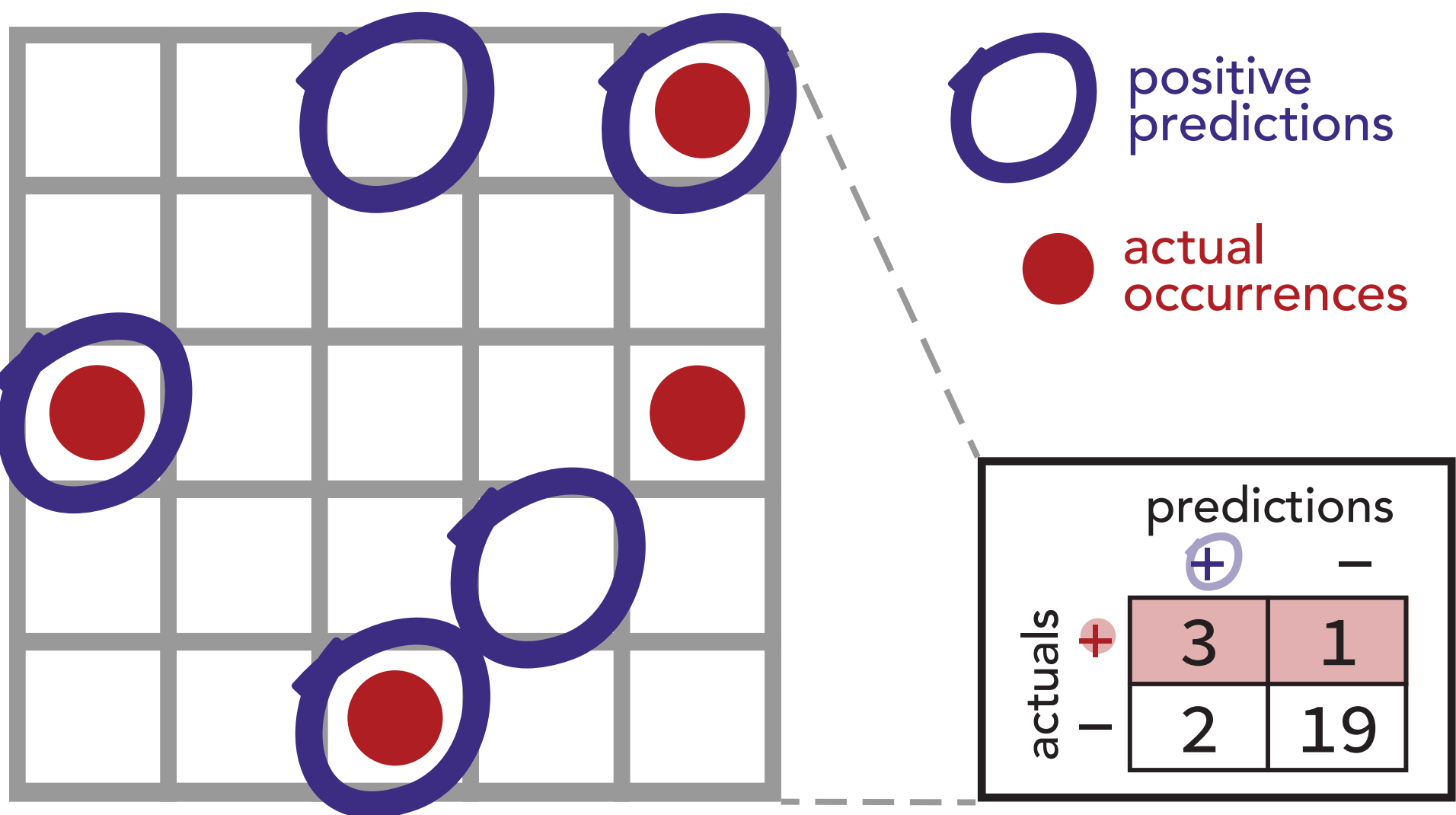


confusion matrix cheat sheet

chrissirico.com



outcomes

True Positives

False Negatives

False Positives

True Negatives

confusion matrix

predictions

+

-

actuals

+

-

TP

FN

FP

TN

Each cell in the grid at far left represents an observation. A hypothetical classification model “circles” observations it thinks are positives. Actual positive occurrences show up as red dots and non-occurrences (negatives) as empty cells. The confusion matrix tallies the model’s correct and incorrect positive and negative predictions (true positives, false negatives, true positives and false positives).

airport security analogy

Your goal is to stop smugglers. Search travelers predicted to be smugglers and pass the rest.

TN: passed, innocent travelers (nice work!)

FP: searched, innocent travelers (oops!)

FN: passed smugglers (oops!)

TP: searched smugglers (nice work!)

classification metrics		
Accuracy <ul style="list-style-type: none">the proportion of all predictions that are correctappears high when dataset is imbalanced, even if model is no better than naïve (always predicts the majority class)		$\frac{TP + TN}{TP + FN + FP + TN}$
Precision (positive predictive value) <ul style="list-style-type: none">accuracy of positive predictionsproportion of searched w/ contraband		$\frac{TP}{TP + FP}$
Recall (sensitivity, true positive rate) <ul style="list-style-type: none">proportion of actual occurrences correctly predicted positiveproportion of smugglers caught		$\frac{TP}{TP + FN}$
Specificity (true negative rate) <ul style="list-style-type: none">proportion of non-occurrences correctly predicted negativeproportion of innocent travelers passed		$\frac{TN}{FP + TN}$
False Positive Rate (false alarm rate) <ul style="list-style-type: none">proportion of non-occurrences falsely predicted positiveproportion of innocent travelers searched		$\frac{FP}{FP + TN}$
False Negative Rate (miss rate) <ul style="list-style-type: none">proportion of actual occurrences falsely predicted negativeproportion of smugglers passed		$\frac{FN}{TP + FN}$
F1 Score <ul style="list-style-type: none">harmonic mean of precision & recallbalances tradeoff between multiple metricsassumes equal value/cost of TP, FP, TN, FN		$2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$

threshold-agnostic metrics vs. binary classification metrics

Threshold-agnostic metrics are based on predicted probabilities. They are useful for describing the strength of model signal across all thresholds and for selecting algorithms, features and hyperparameters. (E.g., log loss, Gini norm, AUC and P-R AUC.)

Classification metrics defined to the left are based on binary (1, 0) predictions. They are useful for selecting a prediction threshold by comparing score tradeoffs resulting from various thresholds.

ROC Curve

AUC (area under the ROC curve)

- threshold-agnostic
- shows (false) lift for naïve models on imbalanced datasets

Precision-Recall Curve

P-R AUC (area under precision-recall curve)

- threshold-agnostic
- better for imbalanced datasets with few positive observations

More machine learning resources at chrissirico.com:

- algorithm, hyperparameter, feature and threshold selection
- training data setup: group/stratified cross validation and time-based targets
- model bias detection, mitigation and fairness metrics

© 2021 Chris Sirico.