confusion matrix

Authors

Alexander Jung*, Department of Computer Science, Aalto University, Espoo, Finland
*Corresponding author: alex.jung@aalto.fi

Abstract

The confusion matrix of a hypothesis on a finite dataset with k label values is the $k \times k$ matrix whose entry (c, c') counts the data points with true label c and prediction c'. Its diagonal counts the correct predictions, and each off-diagonal entry counts one kind of misclassification. The accuracy, the precision, and the recall are read off the matrix by normalizing the diagonal, a column, or a row. On a dataset with label skewness, the confusion matrix exposes failure on the rare class that the accuracy alone hides.

Definition

The weather station in Krems an der Donau (Austria) records the minimum and the maximum air temperature of each day. A linear classifier is learned by logistic regression from 40 such data points from February and April 2024 (Fig. 1). The features of a data point are the two temperatures of one day; its label indicates whether the minimum temperature of the following day stays above 0 °C or falls below it — a frost warning. Most days pose no risk: only 6 of the 40 following days bring frost. The constant prediction "above zero" — a baseline that ignores the features — is therefore correct on 34 of the 40 days, an accuracy of 0.85, and the learned classifier reaches the same accuracy of 0.85. The single number cannot tell the two apart, nor reveal which mistakes either makes: does it miss the frost days, or does it raise false alarms? The confusion matrix breaks the count of predictions down by what was true and what was predicted (Fig. 2): the learned classifier detects one of the six frost days at the cost of one false alarm, while the baseline detects none.

Consider a finite dataset with m data points, each characterized by a feature vector $\mathbf{x}$ and a label y from a finite label space $\mathcal{Y} = \{1, \ldots, k\}$. For a given hypothesis h, the confusion matrix is a $k \times k$ matrix whose row c collects the data points with true label $y = c$ and whose column c' collects those with prediction $h(\mathbf{x}) = c'$ (Hastie et al., 2009): the entry at position (c, c') is the number of data points with $y = c$ and $h(\mathbf{x}) = c'$. The diagonal therefore counts the correctly classified data points, and the off-diagonal entries count each kind of misclassification separately.

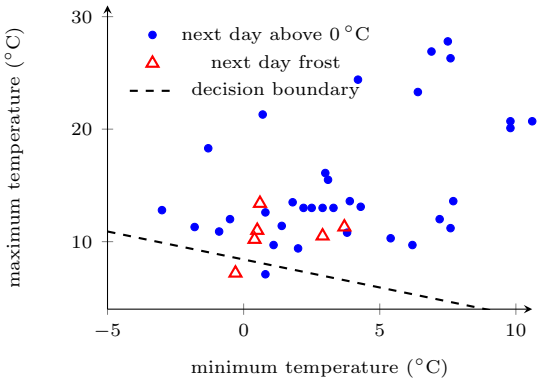

Figure 1: The 40 data points of the Krems frost warning — each day's minimum and maximum temperature, marked by whether the following day stays above 0 °C (filled circles) or brings frost (open triangles) — with the linear decision boundary learned by logistic regression. Data generated by `pythondemos/cm.py`.

learned classifier: accuracy 0.85

true \ predicted	above 0	frost
above 0	33	1
frost	5	1

always-above-zero baseline: accuracy 0.85

true \ predicted	above 0	frost
above 0	34	0
frost	6	0

Figure 2: The confusion matrices of the learned classifier and of the always-above-zero baseline of Fig. 1. Both reach an accuracy of 0.85; only the confusion matrices show that the learned classifier detects one frost day and raises one false alarm, while the baseline detects none. Data generated by `pythondemos/cm.py`.

The standard classification metrics are read off the confusion matrix. The accuracy is the sum of the diagonal divided by m. Normalizing a row by its sum yields the fraction of data points of that class that are detected — for the positive class, the recall — while normalizing a column yields the fraction of predictions of that class that are right, the precision. On a dataset with label skewness, the confusion matrix exposes what a large accuracy can hide: a classifier that never predicts the rare class fills the row of that class with misclassifications while keeping most predictions correct, as the always-above-zero baseline in Fig. 2 does (see baseline).

Cross References

classification, accuracy, precision, recall, label skewness, baseline, label, label space, matrix

References

1. Hastie et al. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed., Springer Science+Business Media, New York, NY, USA.