decision region

Author and editors

Alexander Jung*, Department of Computer Science, Aalto University, Espoo, Finland (ORCID 0000-0001-7538-0990)
Editors: Konstantina Olioumtsevits (ORCID 0000-0003-2697-8659), Ekkehard Schnoor (ORCID 0000-0001-8252-1499), Aalto University, Espoo, Finland
*Corresponding author: alex.jung@aalto.fi

Abstract

A decision region of a classifier is the set of all feature vectors that the classifier maps to one particular label value. The decision regions partition the feature space, one region per label value, and all feature vectors in the same region obtain the same prediction, so the decision regions determine the classifier completely. The union of the boundaries of the decision regions is the decision boundary. The shape of the decision regions depends on the model: a linear classifier has halfspaces, a decision tree has unions of axis-parallel rectangles, and logistic regression on polynomial features has regions separated by a curve.

Definition

A frost warning predicts from the minimum and the maximum air temperature of a day whether the following night brings frost. Any rule that gives this warning splits the pairs of temperatures into two parts: the pairs for which it predicts frost and the pairs for which it does not. Each part is a decision region of the rule. In general, a classifier is a hypothesis $h\colon \mathcal{X} \to \mathcal{Y}$ with a finite label space $\mathcal{Y}$. Its decision region for the label value $\widehat{y} \in \mathcal{Y}$ is the set of feature vectors that it maps to this value,

$$\mathcal{R}_{\widehat{y}} := \big\{\mathbf{x} \in \mathcal{X} : h(\mathbf{x}) = \widehat{y}\big\}.$$

The decision regions partition the feature space: every feature vector lies in exactly one of them, there is one region per label value, and a region may consist of several separate pieces (Bishop, 2006, Sect. 1.5.1). All feature vectors in the same decision region obtain the same prediction, so the decision regions determine the classifier completely. The boundary $\partial\mathcal{R}_{\widehat{y}}$ of a decision region is the set of feature vectors such that every neighborhood contains elements of $\mathcal{R}_{\widehat{y}}$ and of its complement. The union of these boundaries over all label values is the decision boundary.

The shape of the decision regions depends on the model. A linear classifier with $h(\mathbf{x}) = \mathrm{sign}(\mathbf{w}^\top\mathbf{x} + w_0)$ has the two halfspaces on either side of the hyperplane $\mathbf{w}^\top\mathbf{x} + w_0 = 0$ as its decision regions (Bishop, 2006, Sect. 4.1.1). For more than two label values, a linear classifier predicts the value with the largest of K linear scores. Each decision region is then an intersection of $K-1$ halfspaces and therefore convex (Bishop, 2006, Sect. 4.1.2). A decision tree compares

one feature with a threshold at each split, so its decision regions are unions of axis-parallel rectangles (Hastie et al., 2009, Sect. 9.2.1). Logistic regression on polynomial features has decision regions separated by a curve, the level set of a polynomial.

Fig. 1 shows the decision regions of three classifiers learned from the same data points with two features: a linear classifier, a decision tree, and logistic regression on polynomial features. A data point whose label differs from the label value of the region containing its feature vector is misclassified, and the fraction of such data points is the average 0/1 loss. The two label values are separated by a curve and not by any straight line, so the linear classifier misclassifies half of the data points, while the other two misclassify none.

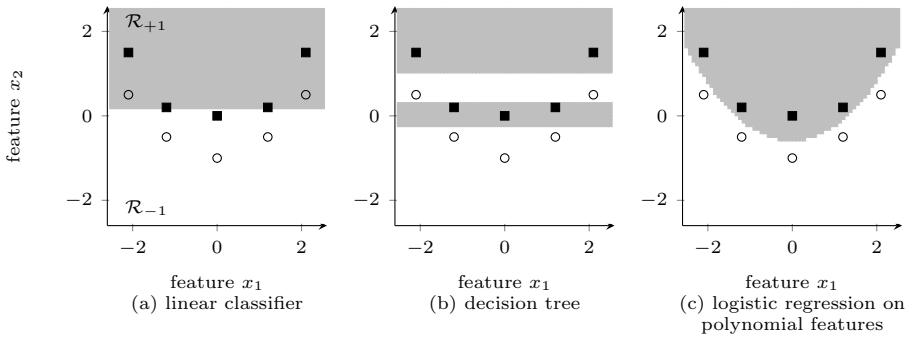


Figure 1: Decision regions of three classifiers learned from the same ten data points with two features. Filled squares are data points with label $y = +1$, and open circles are data points with label $y = -1$. The shaded area is the decision region $\mathcal{R}_{+1}$, the unshaded area the decision region $\mathcal{R}_{-1}$, and the border between them the decision boundary. (a) A linear classifier, whose regions are two halfplanes. (b) A decision tree of depth four, whose regions are unions of axis-parallel rectangles. (c) Logistic regression on all monomials of the two features up to degree two, whose regions are separated by a curve. Data generated by `pythondemos/decisionregion.py`

Cross References

classifier, classification, binary classification, decision boundary, boundary, neighborhood, linear classifier, decision tree, logistic regression, halfspace, hyperplane, feature space, label space, 0/1 loss

References

1. Bishop (2006). Pattern Recognition and Machine Learning. Springer Science+Business Media, New York, NY, USA.

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