Foundations of Machine Learning

Brett Bernstein

August 22, 2018

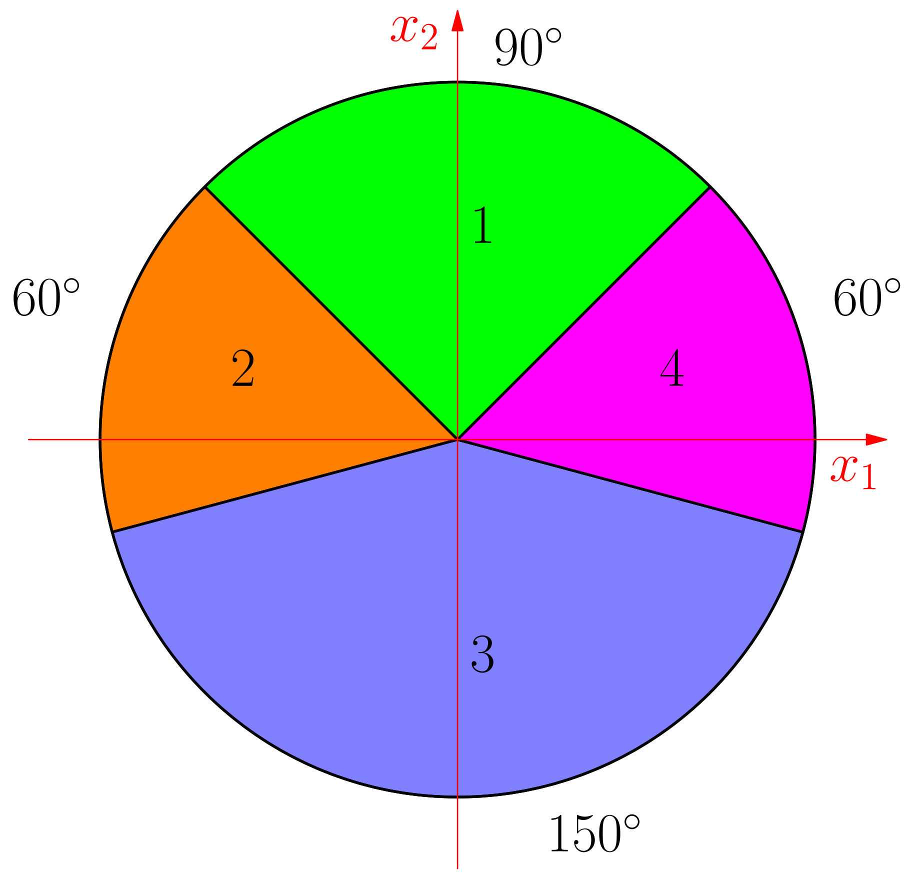
Multiclass: Concept Check

Multiclass Learning Objectives

- Be able to give pseudocode to fit and apply a one-vs-all/one-vs-rest prediction function.
- Be able to describe an example where one-vs-all fails.
- Be able to explain our reframing of multiclass learning in terms of a compatability score function.
- Be able to define the class-specific margin of a data instance using the compatability score function.
- Be able to map a set of linear score functions onto a single linear class-sensitive score function using a class-sensitive feature map. Give some intuition for the value of this feature map (based on features related to the target classes).
- Be able to state the multiclass SVM objective with 1 as the target margin, and be able to generalize using a class-specific target-margin and explain this generalization using the intuition of this target-margin as a lookup table.

Multiclass Concept Check Questions

1. Let $\mathcal{X} = \mathbb{R}^2$ and $\mathcal{Y} = \{1, 2, 3, 4\}$, with X uniformly distributed on $\{x \mid \|x\|_2 \leq 1\}$. Given X, the value of Y is determined according to the following image, where green is 1, orange is 2, blue is 3, and magenta is 4.



For the problems below we are using the 0-1 loss.

(a) Consider the multiclass linear hypothesis space

$$\mathcal{F} = \{f \mid f(x) = \underset{i \in \{1,2,3,4\}}{\arg\max}\, w_i^T x\},$$

where each f is determined by $w_1, w_2, w_3, w_4 \in \mathbb{R}^2$. Give $f_{\mathcal{F}}$, a decision function minimizing the risk over $\mathcal{F}$, by specifying the corresponding w_1, w_2, w_3, w_4. Then give $R(f_{\mathcal{F}})$.

(b) Now consider the restricted hypothesis space

$$\mathcal{F}_1 = \{f \mid f(x) = \underset{i \in \{1,2,3,4\}}{\arg\max}\, w_i^T x, \|w_1\| = \|w_2\| = \|w_3\| = \|w_4\| = 1\}.$$

Consider the decision function $f \in \mathcal{F}_1$ with w_1, w_2, w_3, w_3 set to the angle bisectors of the corresponding regions. Give $R(f)$.

(c) Next consider the class-sensitive version of $\mathcal{F}$:

$$\mathcal{F}_2 = \{f \mid f(x) = \underset{i \in \{1,2,3,4\}}{\arg\max}\, w^T \Psi(x, i)\},$$

where $w \in \mathbb{R}^D$ and $\Psi : \mathbb{R}^2 \times \{1,2,3,4\} \to \mathbb{R}^D$. Give w, Ψ corresponding to $f_{\mathcal{F}_2}$, the decision function minimizing the risk over $\mathcal{F}_2$.

2. Recall that the standard (featurized) SVM objective is given by

$$J_1(w) = \frac{1}{2}\|w\|_2^2 + \frac{C}{n}\sum_{i=1}^{n}[1 - y_i w^T \varphi(x_i)]_+.$$

The 2-class multiclass SVM objective is given by

$$J_2(w) = \frac{1}{2}\|w\|_2^2 + \frac{C}{n}\sum_{i=1}^{n}\max_{y \neq y_i}[1 - m_{i,y}(w)]_+,$$

where $m_{i,y}(w) = w^T\Psi(x_i, y_i) - w^T\Psi(x_i, y)$. Give a Ψ (in terms of φ) so that multiclass with 2 classes $\{-1, +1\}$ is equivalent to our standard SVM objective.

3. Suppose you trained a decision function f from the hypothesis space $\mathcal{F}$ given by

$$\mathcal{F} = \{f \mid f(x) = \arg\max_{i \in \{1,\ldots,k\}} w^T\psi(x, i)\}.$$

Give pseudocode showing how you would use f to forecast the class of a new data point x.

4. Consider a multiclass SVM with objective

$$J(w) = \frac{1}{2}\|w\|_2^2 + \frac{C}{n}\sum_{i=1}^{n}\max_{y \neq y_i}[1 - m_{i,y}(w)]_+,$$

where $m_{i,y}(w) = w^T\Psi(x_i, y_i) - w^T\Psi(x_i, y)$. Assume $\mathcal{Y} = \{1, \ldots, k\}$, $\mathcal{X} = \mathbb{R}^d$, $w \in \mathbb{R}^D$ and $\psi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^D$. Give a kernelized version of the objective.