

Solution 1

Step 1

The general form for a linear predictor is the following:

$$f(x) = w^T x + b$$

We are given three linear functions for each class:

$$\text{Class 1: } w_1 = (1, 1), b_1 = 0 \rightarrow f_1(x) = x_1 \cdot (w_1^1) + x_2 \cdot (w_2^1) + b_1 = x_1 \cdot (1) + x_2 \cdot (1) + 0 = x_1 + x_2$$

$$\text{Class 2: } w_2 = (1, 0), b_2 = 1 \rightarrow f_2(x) = x_1 \cdot (w_1^2) + x_2 \cdot (w_2^2) + b_2 = x_1 \cdot (1) + x_2 \cdot (0) + 1 = x_1 + 1$$

$$\text{Class 3: } w_3 = (0, 1), b_3 = -1 \rightarrow f_3(x) = x_1 \cdot (w_1^3) + x_2 \cdot (w_2^3) + b_3 = x_2 \cdot (0) + x_1 \cdot (1) - 1 = x_2 - 1$$

Hence, for each class, we have the following:

$$f_1(x) = x_1 + x_2,$$

$$f_2(x) = x_1 + 1,$$

$$f_3(x) = x_2 - 1.$$

Step 2

Compare f_1 , f_2 , and f_3 at the intersection points of the decision boundaries:

$$(1 \text{ vs } 2) : f_1 = f_2 \rightarrow x_1 + x_2 = x_1 + 1 \rightarrow x_2 = 1 \rightarrow (0, 1)$$

$$(1 \text{ vs } 3) : f_1 = f_3 \rightarrow x_1 + x_2 = x_2 - 1 \rightarrow x_1 = -1 \rightarrow (-1, 0)$$

$$(2 \text{ vs } 3) : f_2 = f_3 \rightarrow x_1 + 1 = x_2 - 1 \rightarrow x_2 = x_1 + 2 \rightarrow (0, 2) \text{ and } (-2, 0)$$

Hence, we have the following decision boundaries:

$$\text{Decision Boundary between Class 1 and Class 2: } x_2 = 1$$

$$\text{Decision Boundary between Class 1 and Class 3: } x_1 = -1$$

$$\text{Decision Boundary between Class 2 and Class 3: } x_2 = x_1 + 2$$

Step 3

Note: $x_2 = 1$, $x_1 = -1$, and $x_2 = x_1 + 2$ all pass through the common point $(-1, 1)$, and slice the plane into 6 wedges.

Now, select six test points for each wedge, and evaluate f_1 , f_2 , and f_3 to determine the classification.

Wedge label	Test point	$f_1(x)$	$f_2(x)$	$f_3(x)$	Classification
A	$(1, 4)$	5	2	3	Class 1
B	$(1, 2)$	3	2	1	Class 1
C	$(1, -1)$	0	2	-2	Class 2
D	$(-2, -1)$	-4	-1	-3	Class 2
E	$(-3, 0)$	-3	-2	-1	Class 3
F	$(-3, 3)$	0	-2	2	Class 3

Step 4

Visualize points and classify each wedge.

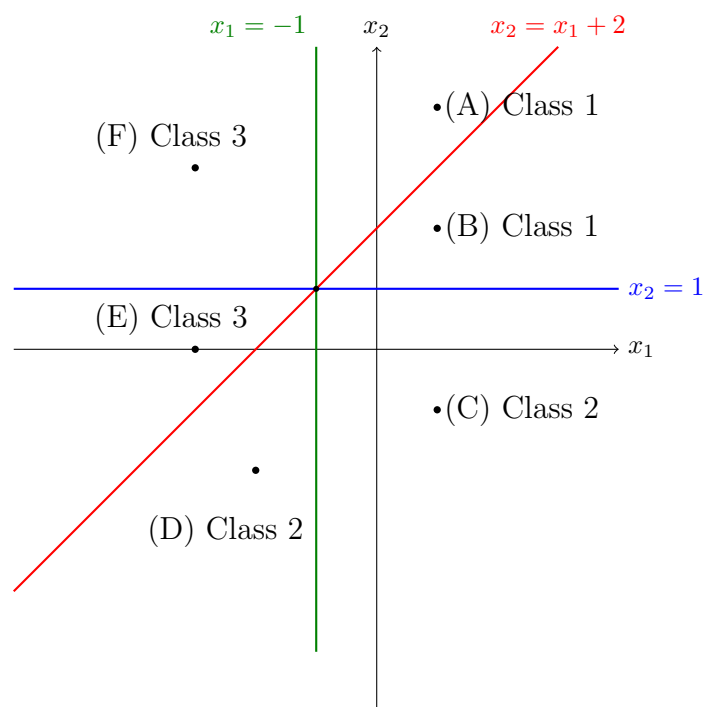


Figure 1: All six wedges produced by the three pairwise boundaries, with the winning class shown in each.

\therefore the decision boundaries and the regions that belong to each class can be seen in *Figure 1*.

Solution 2

Step 1

In order to determine which option will best satisfy the statistical learning framework, we will evaluate each option.

- Option A: A web search on *American plants* will yeild subset of plant images outside of California.
- Option B: Plant photos from a similar region to that of California, will yeild similar flora but not identienal to Californian flora.
- Option C: Plant photos from the favorite Californian city will yeild photos of plants will have a higher bias towards urban Californian flora.

∴ Option B will best satisfy the statistical learning framework, since it will introduce the least amount of bias towards Californian plant identification.

Solution 3

Step 1

We will explore a few potential reasons why the SVM classifier performs well despite the large number of features.

- If the classes the SVM is predicting are sufficiently different, the SVM will be able to find a good decision boundary that separates the data. As a consequence of the difference between the classes, the SVM will have a large margin. This means that the SVM can find a good decision boundary even with a large number of features and minimal training points.
- The generalization error can be expressed as:

$$\text{generalisation error} \approx \sqrt{\frac{R^2/\gamma^2}{n}}$$

So we can see that as the margin grows larger, the generalization error decreases. This means that the SVM can find a good decision boundary even with a large number of features and minimal training points.

\therefore a wide margin for the SVM is a key reason the 1000 training points are sufficient regardless of the one million features in the data.

Solution 4 (a)

The class priors changed: fewer *sports* articles, more *politics*. Within each topic (class), the way the text is written is assumed unchanged, so only the frequency of each label moves.

\therefore this would be an example of a label shift

Solution 4 (b)

The vocabulary has drifted: new proper nouns appear and old ones vanish. This alters the distribution of the features within each topic, while the overall mix of topics (sports, politics, business, etc.) stays the same.

\therefore this is an example of a covariate shift
