CLASSIFICATION

↳ some theory

→ CLASSIFIERS:

↳ You are given some data, with say d features and n samples.

Each observation is in d-dimensional space

↓

You are also given some label that tells you which class a given point belongs to.

↳ Goal: Given some new point, want to predict which class that point belongs to.

→ Example: You are given age, income and credit score as your features, you want to predict whether or not someone will default on their bank loan.

↳ Here our two classes are defaulted and not defaulted.

→ DECISION FUNCTION:

↳ It is a function that takes in a point in the feature space, and spits out a number (scalar) that tells us which class the point belongs to.

$f(x) \geq 0 \rightarrow$ class 1

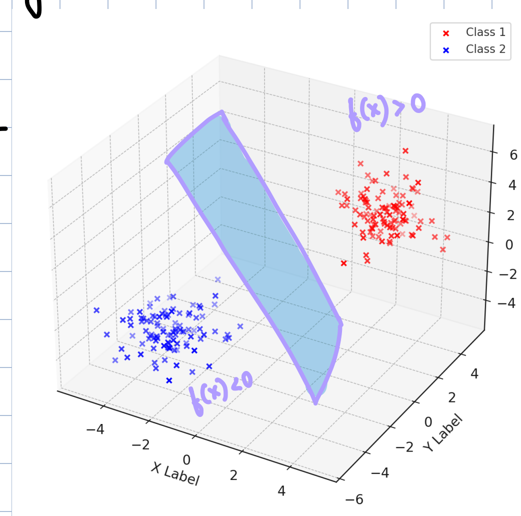
$f(x) < 0 \rightarrow$ class 2

Decision Function

$$f(x) = w \cdot x + \alpha$$

bias

weights



d-1 dimensional surface

→ DECISION BOUNDARY:

$$H = \{x \in \mathbb{R}^d \mid f(x) = 0\}$$

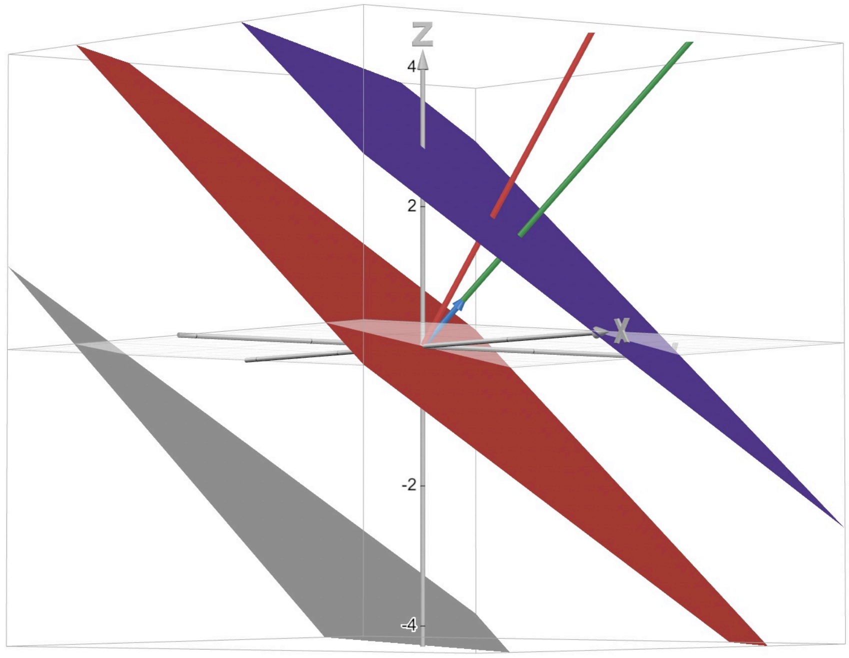
↓

splits the d-dimensional space into two parts.

Decision boundary : $w \cdot x + \alpha = 0$

gives degree of freedom to translate the hyperplane

→ without the alpha term the hyperplane would always pass through the origin (figure below) and this isn't always helpful to fit the data.



purple plane : $\alpha = -20$
red plane : $\alpha = 0$
grey plane : $\alpha = 30$

red vector : some vector in our feature space

Green vector : weight vector.

Blue vector : normalized weight vector.

Note : Finding the most optimal decision boundary is a whole other problem.

SOME IMPORTANT RESULTS :

→ The weight w is orthogonal to the decision boundary :

PROOF :

Our decision boundary is given by $f(x) : w \cdot x + \alpha = 0$

$\Rightarrow w \cdot x = -\alpha$

Now, take two points on the decision boundary say P_1, P_2.

Our vector $x = P_1 - P_2$

$$
\begin{aligned}
w \cdot (P_1 - P_2) &= w \cdot P_1 - w \cdot P_2 \\
&= (-\alpha) - (-\alpha) \\
&= -\alpha + \alpha \\
&= 0
\end{aligned}
$$

→ since the dot product is 0, they are orthogonal.

w is orthogonal to every vector on the hyperplane

⇒ w is orthogonal to the hyperplane.

because of dot product

→ If w is a unit vector, then $f(x) = w \cdot x + \alpha$ is the signed distance from x to H.

positive if on side of w

negative if on other side

EXPLANATION :

↳ Look at the red vector and the blue vector. When you take the dot product between the red and blue vector, you're essentially calculating the distance of the red vector in the direction of the blue vector and the α adjusts for translations (away from the origin)

→ The distance from H to the origin is alpha :

EXPLANATION:

↳ Look at the figure above.