Perceptron

Geometric interpretation

Procheta Sen

UNIVERSITY OF LIVERPOOL

Hyperplane

- The decision in perceptron is made depending on

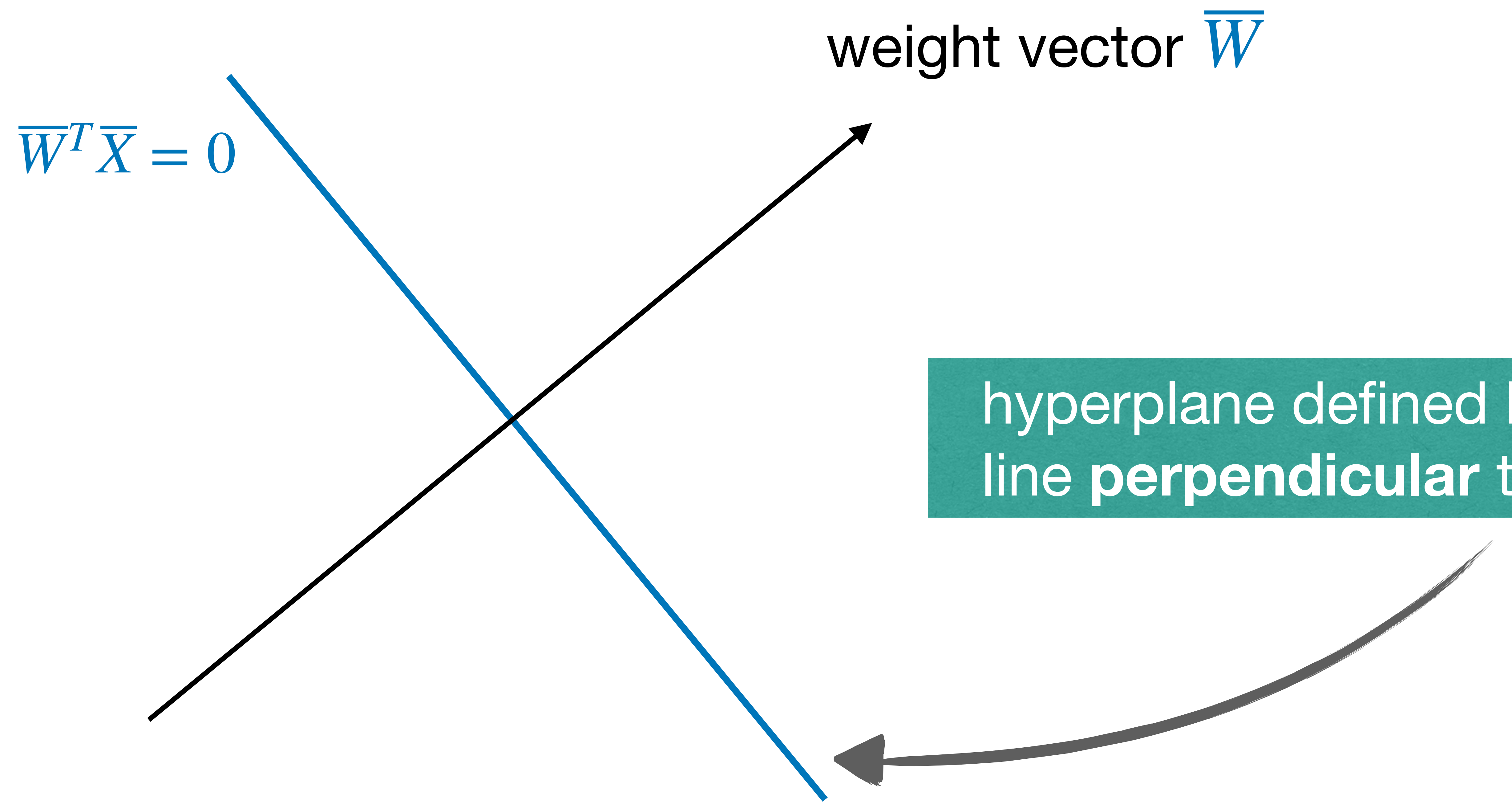
$$\overline{W}^T\overline{X} + b > 0 \quad \text{or} \quad \overline{W}^T\overline{X} + b \leq 0$$

- $\{\overline{X} : \overline{W}^T\overline{X} + b = 0\}$ is the critical region (**decision boundary**)
- $\overline{W}^T\overline{X} + b = 0$ defines a hyperplane

Example:

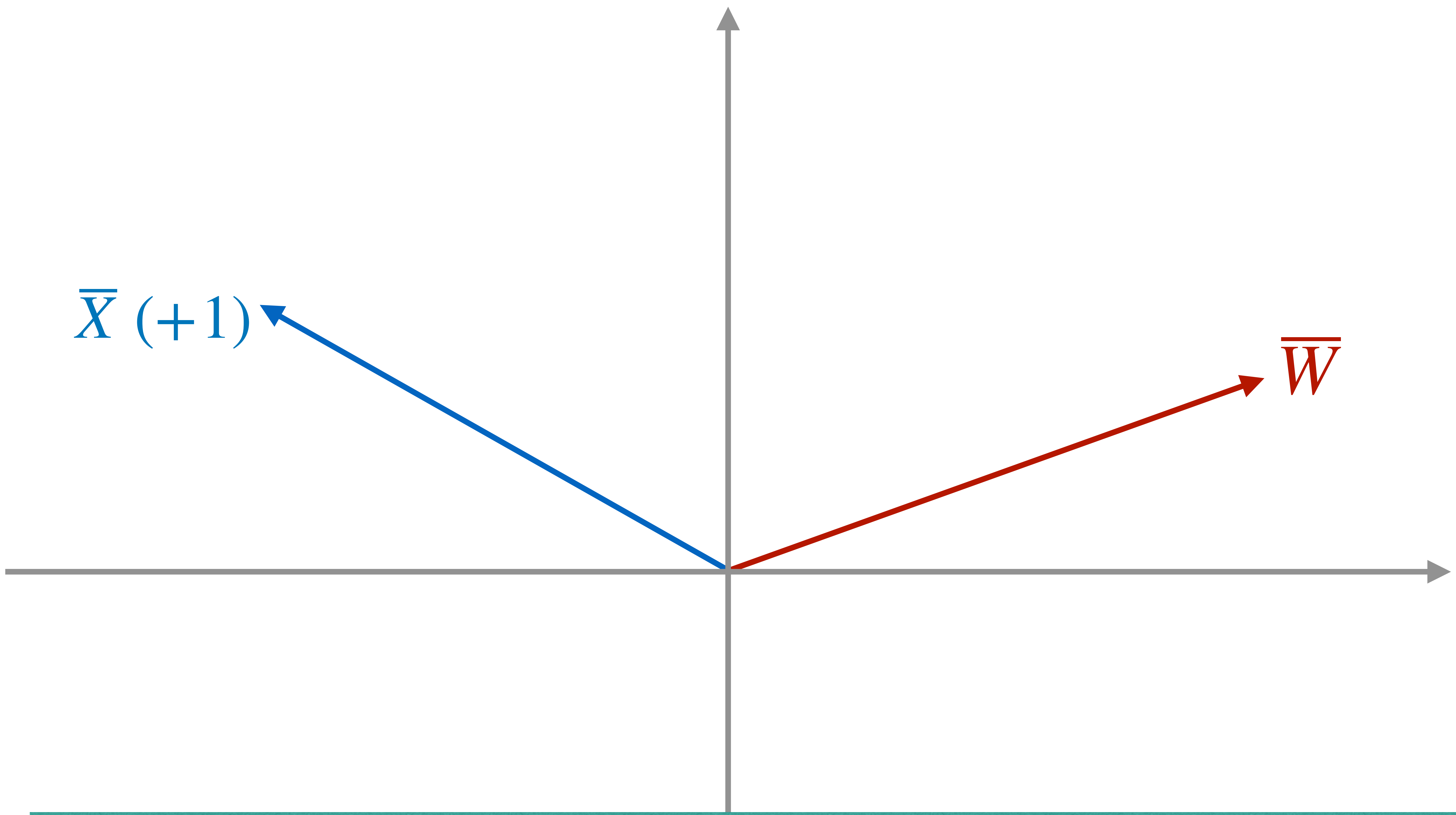
- In 2D space we have $w_1x_1 + w_2x_2 = 0$ (ignoring the bias term), which is a straight line through the origin.
- In N-dimensional space this is an $(N-1)$-dimensional hyperplane.

Geometric interpretation



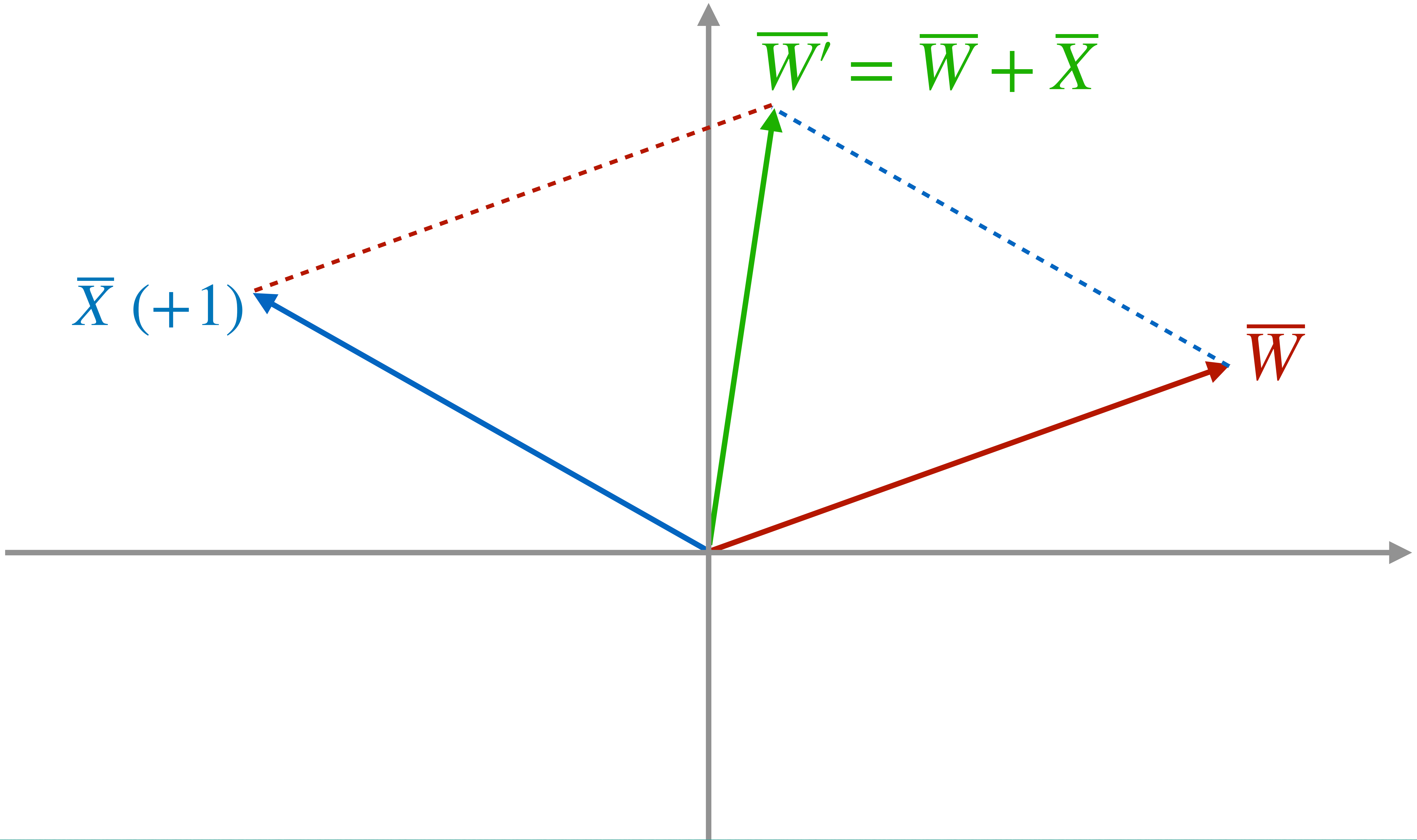
hyperplane defined by the weight vector is the line **perpendicular** to the weight vector

Geometric interpretation



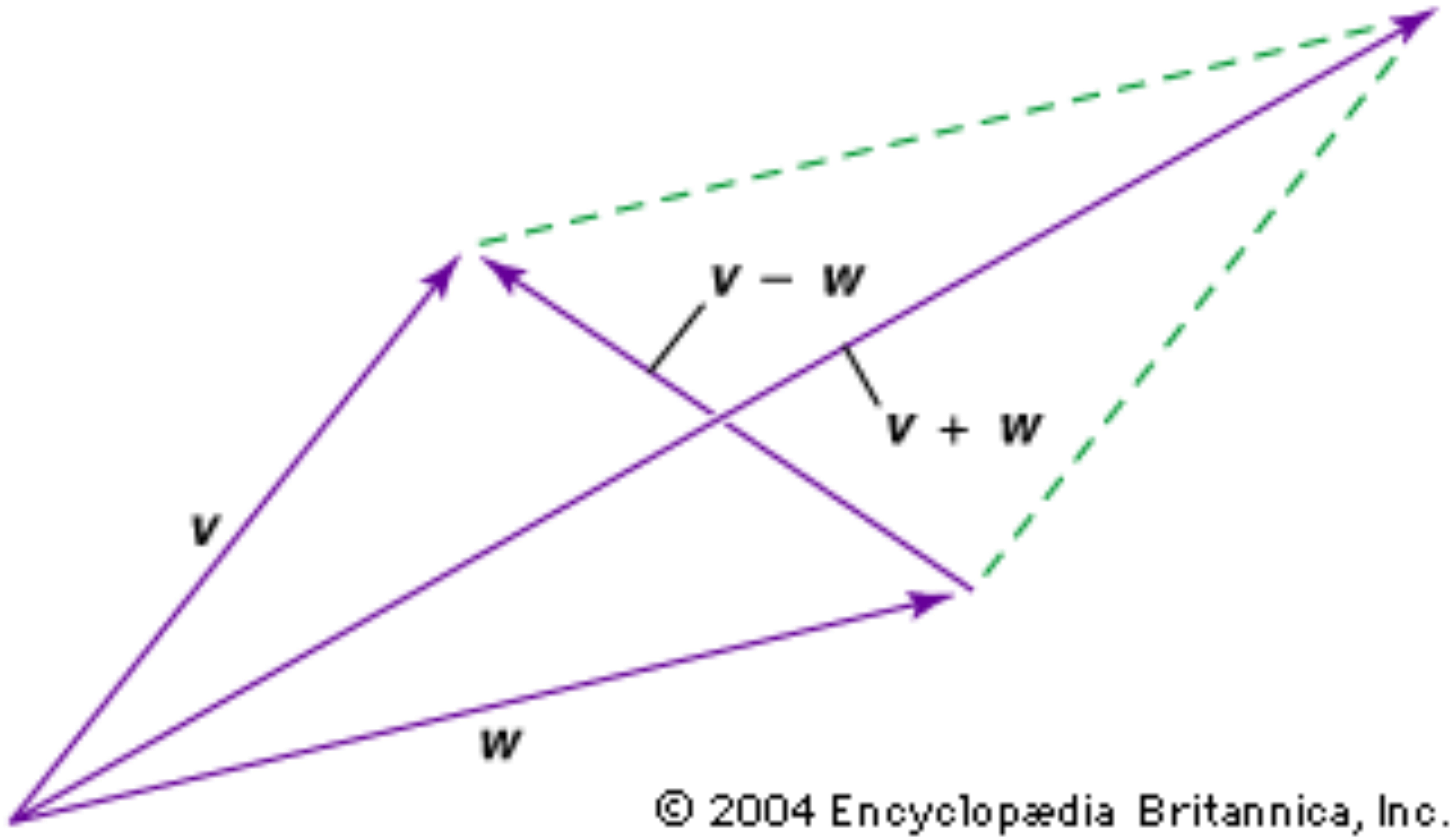
The angle between the current weight vector $\overline{W}$ and the positive instance $\overline{X}$ is greater than 90°. Therefore, $\overline{W}^T\overline{X} < 0$, and this instance is going to get misclassified as negative.

Geometric interpretation



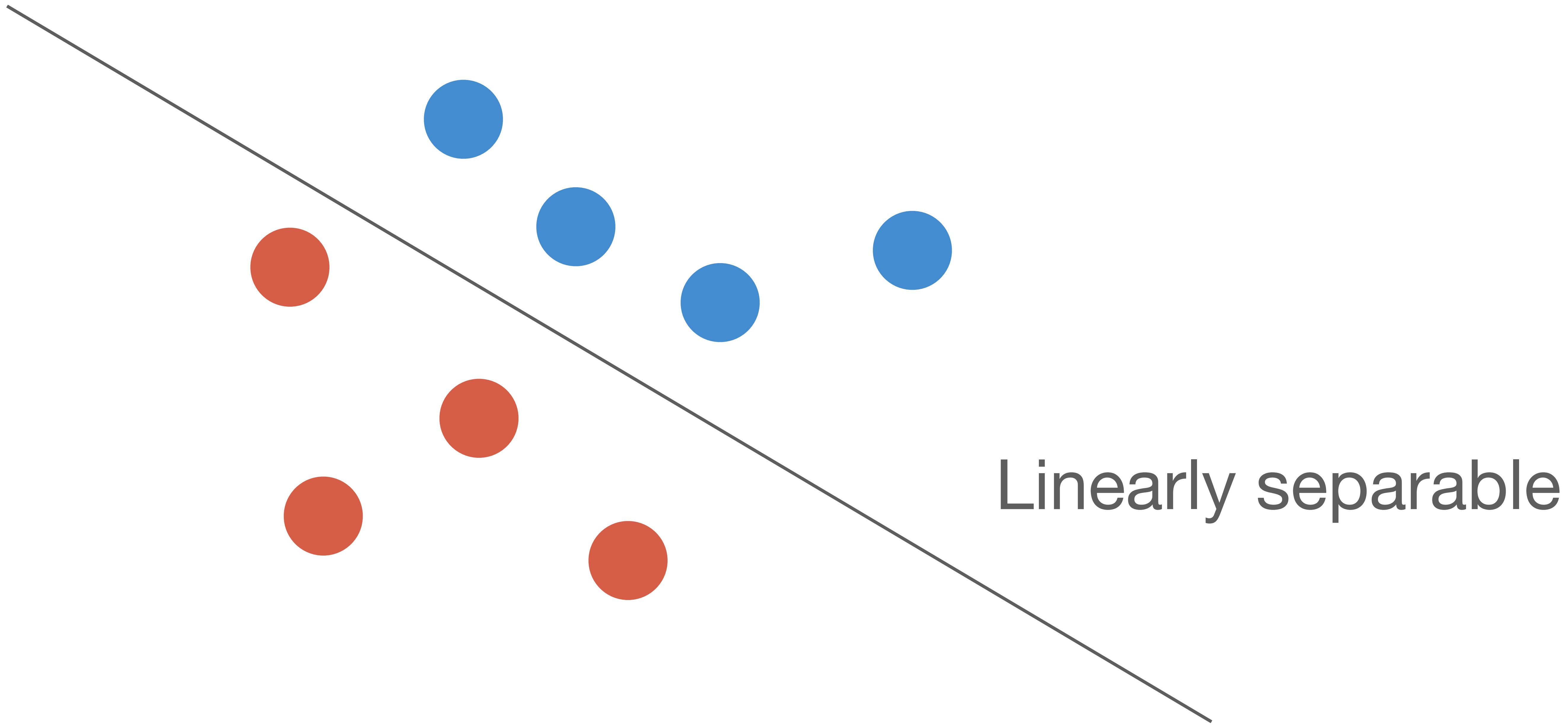
- The new weight vector $\overline{W'}$ is the addition of $\overline{W} + \overline{X}$ (as per the perceptron update rule).
- It lies in between $\overline{X}$ and $\overline{W}$.
- Notice that the angle between $\overline{W'}$ and $\overline{X}$ is less than 90°.
- Therefore, $\overline{X}$ will be classified as positive by $\overline{W'}$.

Vector algebra revision



Linear separability

If a given set of positive and negative training instances can be separated into those two groups using a straight line (hyperplane), then we say that the dataset is **linearly separable**.



Linear separability

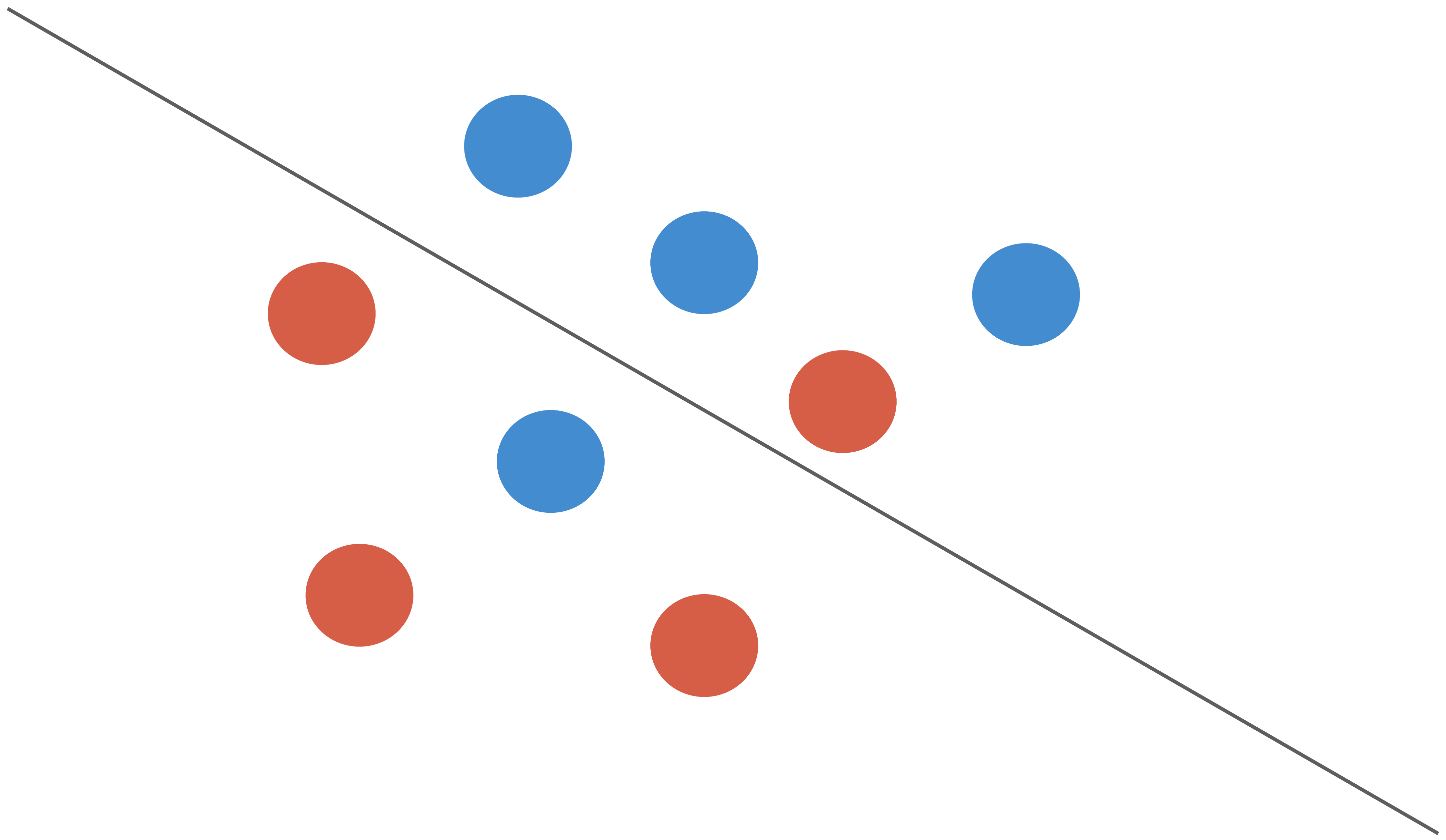
- When a dataset is linearly separable, there can exist more than one hyperplanes that separates the dataset into positive/negative groups.

- In other words, the hyperplane that linearly separates a linearly separable dataset might not be unique.

- However, (by definition) if a dataset is non-linearly separable, then there exist NO hyperplane that separates the dataset into positive/negative groups.

A non-linearly separable case

No matter how we draw straight lines, we cannot separate the red instances from the blue instances



Further remarks

- When a dataset is linearly separable it can be proved that the perceptron will always find a separating hyperplane!

- The final weight vector returned by the Perceptron is more influenced by the final training instances it sees
  - Take the average over all weight vectors during the training (Averaged Perceptron algorithm)