Multiclass Classification

Procheta Sen

UNIVERSITY OF LIVERPOOL

Classification algorithms

Multiclass classifiers

- k-NN
- Naive Bayes

Binary classifiers

- Perceptron
- Logistic regression

How to turn a binary classifier to a multiclass classifier?

Given binary classification algorithm A we want to design a meta-algorithm that use A to make k-class predictions.

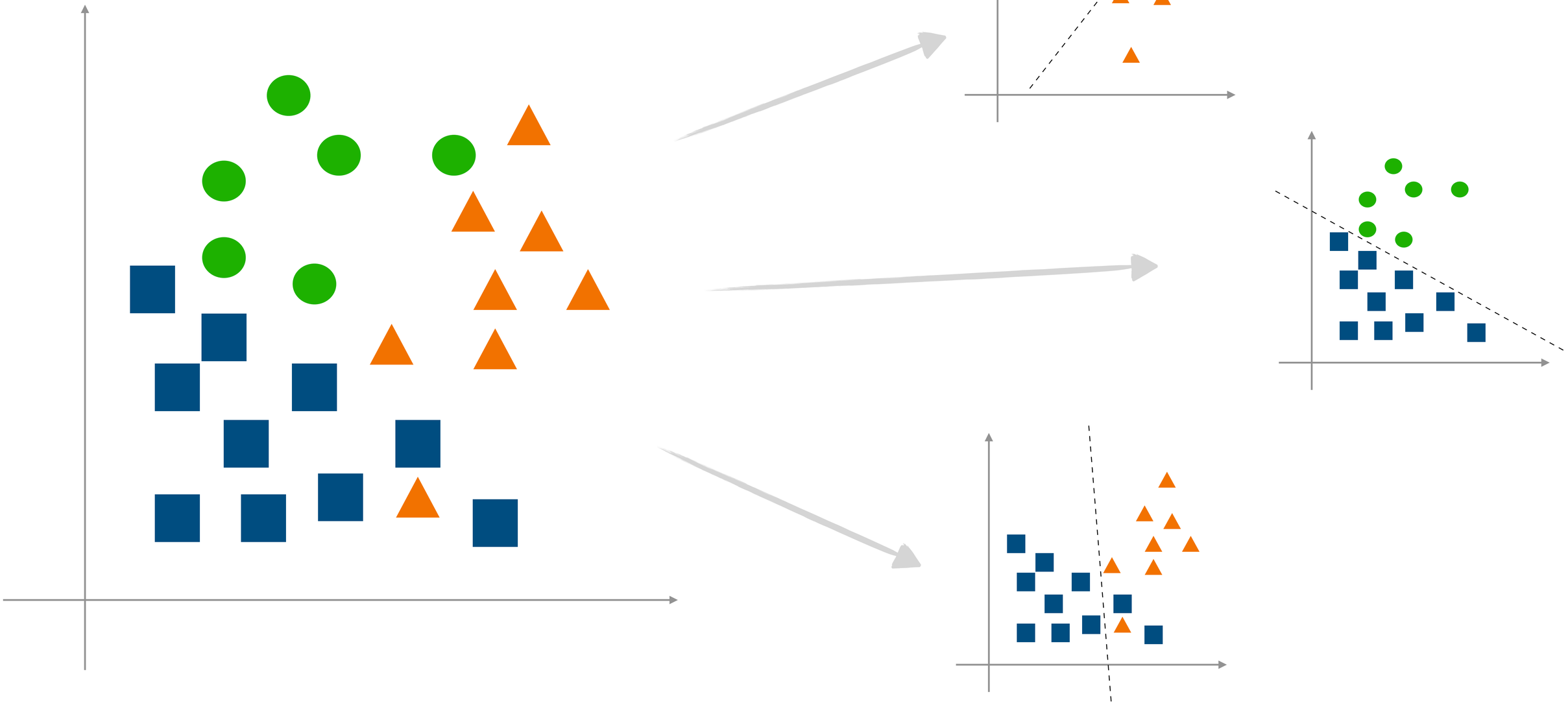
Two strategies

- **One-vs.-one** approach
- **One-vs.-rest** approach

One-vs.-one approach

1. For each pair of classes i and j, use training objects from classes i and j to train algorithm A to distinguish between objects in classes i and j. Denote the obtained classifier $A_{i,j}$.
2. This results in $\frac{k(k-1)}{2}$ prediction models.
3. Applying the prediction model $A_{i,j}$ to an incoming object $\overline{X}$ is interpreted as voting. $A_{i,j}$ votes either +1 for $\overline{X}$ to be in class i, or $A_{i,j}$ votes +1 for $\overline{X}$ to be in class j.
4. For an incoming object $\overline{X}$, apply all prediction models one by one.
5. The class label with the most votes is declared as the winner.

One-vs.-one approach



One-vs.-one approach: drawbacks

- There might be ambiguity if some classes got the same number of votes (if the binary classifier A can produce a confidence score, it can be used to break ties)

One-vs.-rest approach

In this approach we assume that the binary classification algorithm A can output numeric score representing its "confidence" that an object belongs to a particular class.

1. For each class i, train the binary classifier A with the objects of class i treated as positive samples and all other objects as negative samples. Denote the obtained classifier A_i.

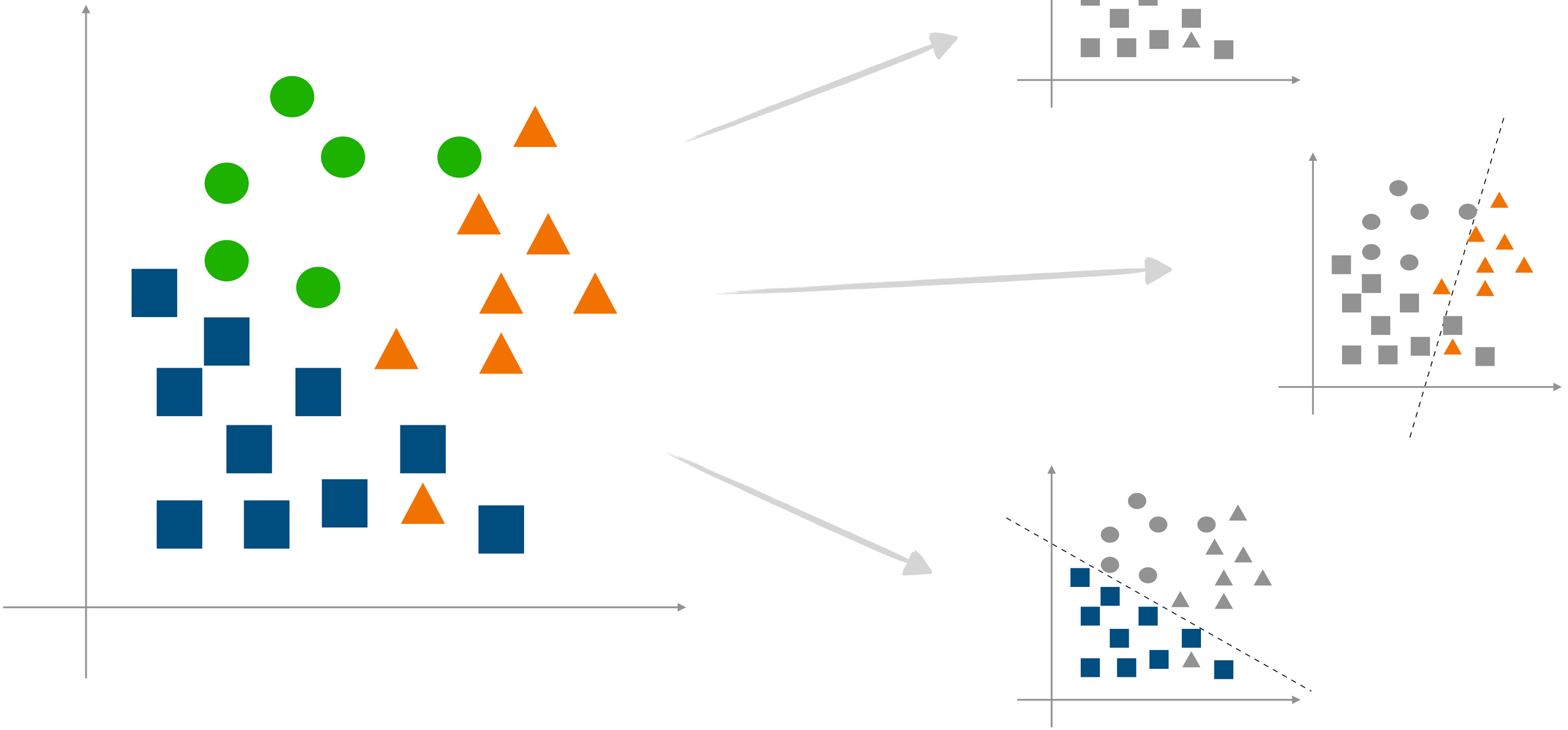
2. This results in k prediction models.

3. For an incoming object $\overline{X}$, apply all prediction models $A_1, A_2, \ldots, A_k$.

4. Output for object $\overline{X}$ the class label y corresponding to the model with the highest score, i.e.

$$y = \text{argmax}_{i \in \{1,2,\ldots,k\}} A_i(\overline{X})$$

One-vs.-rest approach



One-vs.-rest approach

The choice of the numeric score depends on the classifier at hand.

1. For Perceptron: the activation score $a = b + \overline{W}^T \overline{X}$
2. For Logistic regression: $\sigma(a)$, where $a = b + \overline{W}^T \overline{X}$

One-vs.-rest approach: drawbacks

- the scale of the confidence scores may differ between the binary classifiers

- the binary classifiers are trained on unbalanced datasets: usually the set of negative objects will be much larger than the set of positive objects