

Knowledge check

3 minutes

Check your knowledge

1. What does it mean to have an imbalanced dataset?

- ☐ The number of samples is much smaller or greater than is required
- ☐ The feature columns contain many missing values
- ☒ There are many more training examples that correspond to some outputs (categories) than others ✓

Correct. This can lead to the model learning to improve loss metrics by 'cheating.'

2. What information can we extract from a single confusion matrix?

- ☐ Log loss, and/or mean squared error
- ☐ Whether the dataset has overfit the training set
- ☒ What kind of mistakes the model is making ✓

Correct.

3. Why don't we use measures like "True Positives" or "Accuracy" to train our models directly?

- ☐ There are mathematical barriers that prevent these being used for some training regimens
- ☐ Subtle model improvements often do not affect these metrics
- ☒ Both of the above ✓

Correct. In fact, these answers are often related. Methods like gradient descent rely on measures of slope, which do not exist for non-smooth functions like accuracy.

Next unit: Summary

Continue >
