Machine Learning Glossary

Daniel Sanchez

- **Cross validation**: Allows us to compare different machine learning methods and get a sense of how well they will work in practice. It's very common to divide the data into 10 blocks (Ten-Fold Cross Validation).
- **The Confusion Matrix**: The rows correspond to what the `ML` algorithm predicted and the columns to the know truth. The numbers along the main diagonal, tell us how many time the samples were correctly classified; the rest, are samples the algorithm messed up.
- **Sensitivity vs Specificity**: For a $2x2$ matrix:

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}}$$

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

  For bigger matrix, the values in the column, that aren't at the main diagonal, are summed to obtain the False Negatives.
- **Bias vs Variance**:
- **ROC and AUC**:
- **Entropy**:
- **Linear Regression**:
- **Multiple Regression**: