

Classification Review*(find and work with a partner)***1. Evaluation metrics for binary classification**

Investigate the unnormalized confusion matrix below, which is (hypothetical) data for predicted genetic risk of Alzheimer's disease vs. actual incidence of the disease in 100 individuals.

83	7
1	9

- Label the rows with 0 for no disease (first row) and 1 for disease (second row). These are the true labels. Similarly labels the columns with 0 and 1 for the predictions.
- Calculate the *accuracy* of this prediction method.
- Calculate the FPR (*false positive rate*) and TPR (*true positive rate*, also known as *recall*). You can leave these values as fractions.
- Is there anything concerning about these results that is *not* captured by accuracy, FPR, and TPR?

2. Naive Bayes

- If \vec{x} is a vector of *features* and y is the associated *label*, the components of a typical Bayesian model are $p(\vec{x})$, $p(y)$, $p(y|\vec{x})$, $p(\vec{x}|y)$. What are the terms for each of these probabilities?
- Using Bayes' rule, arrange these probabilities into an equation.
- If $\vec{x} = [x_1, x_2]^T$ (two features or $p = 2$), how would you approximate $p(\vec{x}|y)$ using our Naive Bayes assumption?