

where d is the number of features. For $d = 20$, (a modest number of features in a real pattern recognition problem), there would be over one million bins, most of which would necessarily be empty! We see easily that other methods are required. This points up one of the advantages if we can use the variance-covariance matrix Σ . We assume only first-order interactions; there are only d^2 values to estimate, in this case only 400.