

Big deal: extending the application of maximum-likelihood techniques to permit the “learning of parameters governing a distribution from training points.”

- Uncorrupted cases could use $\hat{\vec{\theta}}$ acquired from MLE.
- Iteratively converge on the likelihood for a given data set via Expectation Maximization or Baum-Welch
- Features can be in terms of good features and bad features: $D = \{\vec{x}_1, \dots, \vec{x}_n\}$ or $D = D_g \cup D_b$.