Statistical Learning Methods

Idea given some data (evidence) and have some hypotheses of how the data works, would like to do,
      1)   Choose most likely hypothesis to explain the data
      2)   Predict future data

Example: Bags of candy which can be 1 of 5 types.
      h1:   100% cherry
      h2:   75% cherry 25% lime
      h3:   50% cherry 50% lime
      h4:   25% cherry 75% lime
      h5:   100% lime

$H = \{h_1, \ldots, h_5\}$

Candy chosen from bag revealed one at a time
      $D_1, \ldots, D_n$ Each di is a random variable

Bayesian Learning
      Idea calculate probability of each hypothesis given the data and make a prediction of next item.
D – space of all data sets
d – a vector of observed values from **D**.
      Example: <cherry, cherry, lime, …>

Probability of a given hypothesis
      $p(h_i \mid \mathbf{d}) = \alpha\ p(\mathbf{d} \mid h_i)\, p(h_i)$ (alpha is normalization factor) company might tell us that p(H) = <.1, .2, .4, …>

Say we have p(h_i | d) then calculate probability of next candy as:
      $p(X \mid d) = \Sigma\ \mathbf{P}(X \mid h_i)\, \mathbf{P}(h_i \mid \mathbf{d})$

How to calculate P(**d** | h_i) :
      $P(\mathbf{d} \mid h_i) = \Pi\ P(d_i \mid h_i)$