Contents

0.1 Gaussian Naïve Bayes

Compute the mean and standard deviation to estimate the likelihood.

$$\mu_1 = E[X_1 \mid Y = 1] = \frac{2 + (-1.2) + 2.2}{3} = 1$$

$$\sigma_1^2 = E\left[(X_1 - \mu_1)^2 | Y = 1\right] = \frac{(2-1)^2 + (-1.2-1)^2 + (2.2-1)^2}{3} = 2.43$$

$$P(x_1|Y = 1) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_1-\mu_1)^2}{2\sigma^2}} = \frac{1}{3.91} e^{-\frac{(x_1-1)^2}{4.86}}$$

0.2 Bayesian Belief Network

- Naïve Bayes classifier works with the assumption that the values of the input features are conditionally independent given the target value.
- This assumption dramatically reduces the complexity of learning the target function.
- Bayesian Belief Network describes the probability distribution governing a set of variables by specifying a set of conditional independence assumptions along with a set of conditional probabilities. Conditional independence assumptions here apply to subsets of the variables.

$$P(x_1, x_2, \ldots, x_l \mid {x_1}', {x_2}`, \ldots, {x_m}`, y_1, y_2, \ldots, y_n) = P(x_1, x_2, \ldots, x_l | y_1, y_2, \ldots, y_n)$$

0.3 Training Bayesian Classifier

During training, typically log-space is used.

$$\begin{aligned} y_{NB} &= \arg\max_y \left[\log P(y)\Pi_{i=1}^{n} P(x_i|y)\right] \\ &= \arg\max_y \left[\log P(y) + \sum_{i=1}^{n} \log P(x_i|y)\right] \end{aligned}$$

0.4 Text Classification

Algorithm 0.1 Text-based Naïve Bayes Classification

function TRAIN-NAIVE-BAYES(D, C) **returns** $\log P(c)$ and $\log P(w|c)$
  for all class $c \in C$ **do** ▷ Calculate $P(c)$ terms
    $N_{doc} \leftarrow$ number of documents in D
    $N_c \leftarrow$ number of documents from D in class c
    $logprior[c] \leftarrow \log \frac{N_c}{N_{doc}}$
    $V \leftarrow$ vocabulary of D
    $bigdoc[c] \leftarrow$ APPEND(d) **for** $d \in D$ **with** class c
    for all word w in V **do** ▷ Calculate $P(w|c)$ terms
      COUNT$(w, c) \ldots$
    end for
  end for
end function

The word with doesn't occur in the training set, so we drop it completely (we don't use unknown word models for Naïve Bayes)

0.5 Evaluating Classifiers

- Gold Label is the correct output class label of an input.
- Confusion Matrix is a table for visualizing how a classifier performs with respect to the fold labels, using two dimensions (system output and gold labels), and each cell labeling a set of possible outcomes.
- True Positives and True Negatives are correctly classified outputs belonging to the positive and negative class, respectively.

0.6 Precision, Recall, F-Measure

$$\textbf{Precision} = \frac{\text{true positives}}{\text{true positives } + \text{ false positives}} \tag{1}$$

0.7 ROC Curve

- A receiver operating characteristic curve (ROC curve) is a graphical plot that illustrates the performance of a binary classifier model.
- The ROC curve is the plot of the true positive rate (recall) (TPR) against the false positive rate (FPR).
- ROC curve plots TPR vs. FPR at different classification thresholds.
- Classification threshold is used to convert the output of a probabilistic classifier into class labels.
- The threshold determines the