Week 4

Lecture

- Bayes Theorem
  - Prior, likelihood, posterior
  - Naïve Bayes
- Evaluation Methods
  - Holdout method/ cross validation/ leave-one-out cv/ cv for parameter tuning
- Performance measures
  - Accuracy, recall, precision and f1 score, confusion matrix

Tutorial

Task: Implementing NBC for classifying iris data.

Step 1: Load the training/test data.
Step 2: Create Gaussian NB from sklearn
Step 3: Cross validation for evaluating performance
Step 4: Grid Search with cv for parameter selection

TODO:

1. Implement GaussianNB.
2. Calculate precision, recall, F1 and confusion matrix and answer the questions.
3. Compare NB's cv accuracy with train/test split accuracy
4. Advantages and disadvantages of LOOCV?

Naïve Bayes

Week 4 Lecture Example

Assumption: the features are conditionally independent given the class label.

$p(\mathbf{x} \mid y = c, \theta) = \prod_{j=1}^{D} p(x_j \mid y = c, \theta_{jc})$

The form of class-conditional density depends on the type of each feature.

- Real-valued features

  $$p(\mathbf{x} \mid y = c, \theta) = \prod_{j=1}^{D} \mathcal{N}(x_j \mid \mu_{jc}, \sigma_{jc}^2) = \prod_{j=1}^{D} \{ \frac{1}{\sigma_{jc}\sqrt{2\pi}} \exp^{-\frac{(x_j - \mu_{jc})^2}{2\sigma_{jc}^2}} \}$$

- Binary features ($x_j \in \{0, 1\}$)

  $p(\mathbf{x} \mid y = c, \theta) = \prod_{j=1}^{D} \mathrm{Ber}(x_j \mid p_{jc}) = \prod_{j=1}^{D} \{ p_{jc}^{x_j} (1 - p_{jc})^{1 - x_j} \}$

- Categorical features ($x_j \in \{1, \ldots, K\}$)

  $p(\mathbf{x} \mid y = c, \theta) = \prod_{j=1}^{D} \mathrm{Cat}(x_j \mid \mu_{jc})$ where μ_{jc} is a histogram over the possible values for x_j in class c.

sum rule $\quad p(X) = \sum_{Y} p(X, Y) \qquad p(x) = \int p(x, y)\, \mathrm{d}y$

product rule $\quad p(X, Y) = p(Y|X)p(X). \qquad p(x, y) = p(y|x)p(x).$

Bayes Rule

$$p(X, Y) = p(Y|X)p(X).$$

$$\downarrow$$

$$p(X, Y) = p(Y, X)$$

$$\downarrow$$

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$$

$$\downarrow$$

$$p(X) = \sum_{Y} p(X|Y)p(Y).$$

$$p(\mathbf{w}|\mathcal{D}) = \frac{p(\mathcal{D}|\mathbf{w})p(\mathbf{w})}{p(\mathcal{D})} \qquad p(\mathcal{D}) = \int p(\mathcal{D}|\mathbf{w})p(\mathbf{w})\, \mathrm{d}\mathbf{w}.$$

posterior $\propto$ likelihood $\times$ prior