TDA 231 Machine Learning: Homework 2

Goal: Classification
Grader: Jonatan Kilhamn

Due Date: February 6, 2017

General guidelines:

1. All solutions to theoretical problems, and discussion regarding practical problems, should be submitted in a single file named *report.pdf*
2. All matlab files have to be submitted as a single zip file named *code.zip*.
3. The report should clearly indicate your name, personal number and email address
4. All datasets can be downloaded from the course website.
5. All plots, tables and additional information should be included in *report.pdf*

1 Theoretical problems

Problem 1.1 [Bayes classifier, 6 points]

A psychologist does a small survey on "happiness". Each respondent provides a vector with entries 1 or 0 corresponding to if they answered "yes" or "no" to a question respectively. The question vector has attributes

$$\mathbf{x} = (\text{rich, married, healthy}) \tag{1}$$

Thus a response $(1, 0, 1)$ would indicate that the respondent was "rich", "unmarried" and "healthy". In addition, each respondent gives a value $c = 1$ if they are content wih their life and $c = 0$ if they're not. The following responses were obtained.

$$\begin{aligned} c &= 1 : (1,1,1), (0,0,1), (1,1,0)(1,0,1) \\ c &= 0 : (0,0,0), (1,0,0), (0,0,1), (0,1,0) \end{aligned}$$

(a) Using naive Bayes, what is the probability that a person who is ("not rich", "married" and "healthy") is "content"?

(b) What is the probability that a person who is "not rich" and "married" is content (i.e. we do not know if they are "healthy").

Problem 1.2 [Extending naive Bayes, 4 points]

Consider now, the following vector of attributes:

1. $x_1 = 1$ if customer is younger than 20 and 0 otherwise.
2. $x_2 = 1$ if customer is between 20 and 30 in age, and 0 otherwise.
3. $x_3 = 1$ if customer is older than 30 and 0 otherwise
4. $x_4 = 1$ if customer walks to work and 0 otherwise.

Each vector of attributes has a label "rich" or "poor". Point out potential difficulties with your approach above to training using naive Bayes. Suggest how to extend your naive Bayes method to this dataset.

2 Practical problems

Problem 2.1 [Bayes classifier, 5 points]

Dowload the dataset `dataset2.mat`. The dataset contains 3-dimensional data, X, generated from 2 classes with labels, $\mathbf{y}$ either $+1$ or -1. Each row of X and $\mathbf{y}$ contain one observation and one label respectively. There are 1000 instances of each class.

(a) Assume that the class conditional density is spherical Gaussian, and both classes have equal prior. Write the expression for the Bayes (not naive-Bayes!) classifier i.e. derive

$$P(y_{new} = -1 | \mathbf{x}_{new}, X, \mathbf{y}) \quad (2)$$

$$P(y_{new} = +1 | \mathbf{x}_{new}, X, \mathbf{y}) \ . \quad (3)$$

It is useful to note that the dependence on training data $X, \mathbf{y}$ for class 1 can be expressed as:

$$P(\mathbf{x}_{new} | y_{new} = 1, X, \mathbf{y}) = P(\mathbf{x}_{new} | \hat{\boldsymbol{\mu}}_1, \hat{\sigma}_1^2)$$

where $\hat{\boldsymbol{\mu}}_1 \in \mathbb{R}^3$ and $\hat{\sigma}_1^2 \in \mathbb{R}$ are MLE estimates for mean (3-dimensional) and variance based on training data with label $+1$ (and similarly for class 2 with label -1).

(b) Implement a function `sph_bayes()` which computes the probability of a new test point `Xtest` coming from class 1 (`P1`) and class 2 (`P2`). Finally, assign a label `Ytest` to the test point based on the probabilities `P1` and `P2`.

```
function [P1, P2, Ytest]=sph_bayes(Xtest, ...) % other parameters needed.
```

(c) Write a function

```
function [Ytest] = new_classifier(Xtest, mu1, mu2)
```

which implements the following classifier, we call *new_classifier*,

$$f(\mathbf{x}) = \text{sign}\left(\frac{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^\top (\mathbf{x} - \mathbf{b})}{\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|_2}\right)$$

The parameter $\mathbf{b} = \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)$.

(d) Report 5-fold cross validation error for both classifiers, using the ML estimates for μ.

Problem 2.2 [Handwritten digit recognition, 5p]

Download `digits.mat`. This dataset contains a $256 \times 1100 \times 10$ matrix *data* comprising 1100 images of handwritten digits $(1 - 9, 0)$ each of size 16×16. For example $data(:, 45, 3)$ is the feature vector (data) corresponding to 45^{th} sample for class 3 (corresponding to digit 3). You can visualize this using the commands

```
y = reshape(data(:, 45, 3) , 16, 16); % 16x16 image
imshow(y);                            % show image
```

(a) Use `new_classifier` designed previously to do binary classification between class 5 and 8.

(b) Investigate an alternative feature function as described below:

1. Scale each pixel value to range $[0, 1]$ from original gray-scale $(0 - 255)$.
2. Compute variance of each row and column of the image. This will give you a new feature vector of size 32 i.e.

$$\mathbf{x}' = [\ \mathrm{Var}(row_1), \mathrm{Var}(row_2), \ldots, \mathrm{Var}(row_{16}), \mathrm{Var}(col_1), \ldots, \mathrm{Var}(col_{16})\ ]^T$$

For example, the variance of the scaled first row of image can be obtained as:

```
z = y(1, :)/255;          % scaled row
var_scaled = var(z);      % variance
```

(c) Report 5-fold cross validation results for parts (a) and (b) in a single table. What can you say about the results?

(d) (Optional !!!) What do you think is the easiest thing that can improve above results? Did you try it and get any improvement?