CMPS 242: Machine Learning: Fall 2016: HW 4

Due: 22nd November 2016

General Instructions (please review carefully):

- The assignment is to be attempted in groups of two. If you choose to not work with a partner, **two** points will be automatically deducted from your score for this homework. It is your responsibility to find a partner.
- Each group needs to submit only one set of solutions. However, each group member should completely understand the group's solutions. It is **strongly** recommended that each student work on all of the problems individually before integrating their solutions into the group consensus. Dividing up the problems so each student does just a subset of them is the wrong approach.
- LaTeX is preferred, but neatly handwritten solutions will also be accepted. All solutions need to be handed over in the class before the beginning of the lectures.
- The names of the group members, and their UCSC ID (@ucsc.edu email address) should prominently be written on the upper left corner of the first page.
- Multiple sheets should be stapled together in the upper **left** corner.
- Solutions to the problems should be clearly labeled with the problem number.
- Although no points are given for neatness, illegible and/or poorly organized solutions can be penalized at the TA discretion.
- Clearly acknowledge sources (web, people, books, etc.), and mention if you discussed the problems with other students or groups. In all cases, the course policy on collaboration applies, and you should refrain from getting direct answers from anybody or any source. If in doubt, please ask the instructors or TAs.

Question 1 (4 points): For this assignment generate the training dataset as follows: draw 1000 points from the 2-d Gaussian distribution with $\mu_+ = (1, 1)$, $\Sigma_+ = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and label them as $+1$. Similarly, draw another 1000 points from the 2-d Gaussian distribution with $\mu_- = (-1, -1)$, $\Sigma_- = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$ and label them as -1. From the same distributions as above draw 500 points from each class for the test dataset.

- Download and install LibSVM from `http://www.csie.ntu.edu.tw/~cjlin/libsvm/`.
- Set $C = 1$ and train Support Vector Machines (SVM) with Gaussian Kernels whose kernel widths are given by $\gamma = \{1, 10, 100, 1000\}$
- Plot the decision boundary for each value of γ. Also report the test accuracy. Comment on what you observe.
- Fix $\gamma = 10$ and let $C \in \{1, 10, 100, 1000\}$. As before plot the decision boundary and report the test accuracy. What do you observe?

Question 2 (2 points): Consider a dataset $\{(\mathbf{x}_n, t_n)\}$ for $n = 1, \ldots, N$, where $\mathbf{x}_n \in \mathbb{R}^D$. For each of the statements below, answer either True or False. In either case justify your answer for full credit. In some cases, the answer can be neither True nor False.

- In the case of the linear SVM, it is always preferable to solve the Primal optimization problem.
- If $N \gg D$ then it is preferable to solve the Linear SVM problem in the dual.
- If we use the feature map $\phi(\mathbf{x}) = \begin{bmatrix} \mathbf{x} \\ \mathbf{x}^2 \end{bmatrix}$, where $\mathbf{x}^2$ is a vector given by squaring every entry of $\mathbf{x}$, then the primal is a $2D$ dimensional optimization problem.
- In the above case, the dual is a $2N$ dimensional optimization problem.

Question 3 (2 points): For each of the datasets below, indicate if one should use a CRF for classification. Justify your answer

- Scanned OCR digits, where the task is to predict the label $\in \{0, \ldots, 9\}$
- Stock price of AMZN, where the task is to predict the value of the stock in the future
- The set of movies that a user has watched, where the task is to predict the next movie that the user is going to watch
- News articles from NYTimes, where the task is to predict if a word is a proper noun.

Question 4 (2 points): Consider a Hidden Markov Model where the the hidden state y takes on two values namely 1 and 2, with the corresponding state transition matrix given by $M = \begin{bmatrix} 0.3 & 0.7 \\ 0.1 & 0.9 \end{bmatrix}$. Moreover, let the observed state x take on values 0 or 1. The emission probabilities are given by

$$p\,(x = 0|y = 1) = 0.1 \text{ and } p\,(x = 1|x = 1) = 0.9$$
$$p\,(x = 0|y = 1) = 0.5 \text{ and } p\,(x = 1|x = 1) = 0.5$$

Use the forward-backward algorithm to compute the probability of observing the sequence 0101. You may assume that at the beginning both the hidden states have equal probability.

Learning Outcomes After this homework you should

- understand how the choice of kernel width and regularization parameter affect the decision boundary of a SVM
- gain a deeper understanding of the connections between the primal and dual optimization problems for SVM, and their properties
- be able to infer classification tasks where CRFs/HMMs are applicable
- understand the forward-backwards algorithm used for inference in HMMs.

Changelog

- 09 November 2016: First version created.