Advanced Techniques in Machine Learning (89-654-01)
Exam - Moed Aleph

Dr. Yossi Keshet
July 27, 2016

Instructions
- Total time: 3 hours.
- All written or printed material is allowed.
- You should answer all questions.
- Explain your steps.
- Good luck!

1. A modified version of the Perceptron algorithm is given in Figure 1.

   INIT: training set $S = \{(x_i, y_i)\}_{i=1}^m$
   INITIALIZE: $w = 0$
   LOOP:
   - choose example $(x_i, y_i)$ uniformly at random from $S$
   - if $y_i w \cdot x_i < -1$:
     - update: $w = w + y_i x_i$

   Figure 1: Large margin Perceptron

(a) Assume that $\|x_i\| \leq R$ for all $i$ and that there exists a vector $u$, $\|u\| = 1$ such that $y_i (u \cdot x_i) \geq \gamma$ for all $i$. Derive an upper bound for the number of mistakes made by this new Perceptron algorithm. [15 pt]

(b) How the new bound you derived in (a) compares to the standard Perceptron bound and why? [4 pt]

(c) Write the new Perceptron algorithm in its kernel from? That is, instead of $w$ use its implicit definition:

\[ w \cdot x = \sum_{j=1}^{i} \alpha_j x_j \cdot x \]

What are $\alpha_j$ in this algorithm? [8 pt]

(d) What is the advantage of the negative margin Perceptron algorithm when working with kernels and huge amount of training data [7 pt]
2. You would like to write a software for solving crossword puzzles (TASHBETZ). In order to do so, the user of the software inputs in the a search bar terms and the tool returns a lift of words. For each term \( t \in T \) (\( T \) is the set of known terms) the system pulls a set of \( N \) candidate word definitions. Each candidate definition has a set of features \( x \in \mathbb{R}^d \) and a label \( y \in \{-1, +1\} \) indicating if the definition is relevant or not.

(a) Given a training set \( S \) of \( m \) terms with \( N \) candidate definitions,

\[
S = \{(q^i, (x^i_1, \ldots, x^i_N), y^i)\}_{i=1}^m,
\]

propose an algorithm for the task. [28 pt]

(b) You would like to allow users of the tools to add their own definitions. Do you need to retrain you algorithm every time a new definition is entered? explain. [5 pt]

3. We would like to predict where is a square located in an image as in Figure 2. Each square is defined by a vector of four points \( y = (y_1, y_2, y_3, y_4) \). The predicted square is denoted by \( \hat{y} = (\hat{y}_1, \hat{y}_2, \hat{y}_3, \hat{y}_4) \). The performance are measure by the intersection between the squares \( y \) and \( \hat{y} \) divided by their union (called intersection-over-union), and denoted \( \gamma(y, \hat{y}) \).

(a) Propose an algorithm that predicts \( \hat{y} \) and aims at maximizing \( \gamma(y, \hat{y}) \) given a training set of images and labeled squares. [28 pt]

(b) Suggest a feature function for such a prediction. The feature should not be describe mathematically, just conceptually with words. [5 pt]

![Figure 2: Image with a square](image-url)