Machine Learning
Homework 3

Due on May 24, 2017

1. Let

$$\mathbf{g}(\mathbf{x}) = \begin{bmatrix} g_1(\mathbf{x}) \\ g_2(\mathbf{x}) \\ \vdots \\ g_m(\mathbf{x}) \end{bmatrix} \quad \text{and} \quad \mathbf{h}(\mathbf{x}) = \begin{bmatrix} h_1(\mathbf{x}) \\ h_2(\mathbf{x}) \\ \vdots \\ h_k(\mathbf{x}) \end{bmatrix}$$

   where $g_i : R^n \to R$ be a *convex* function for all $i = 1, 2 \ldots, m$, and $h_j : R^n \to R$ be a *linear* function for all $j = 1, 2 \ldots, k$.

   Consider $\mathcal{F} = \{\mathbf{x} \mid \mathbf{g}(\mathbf{x}) \leq \mathbf{0},\ \mathbf{h}(\mathbf{x}) = \mathbf{0}\} \subset R^n$. Prove that $\mathcal{F}$ is a *convex* set. (15 %)

2. Prove that for any matrix $B \in R^{m \times n}$, either the system (**I**)

$$B\mathbf{x} < \mathbf{0}$$

   or the system (**II**)

$$B^\top \alpha = \mathbf{0}\ , \alpha \geq \mathbf{0} \text{ and } \alpha \neq \mathbf{0}$$

   has a solution but *never both.* (15 %)
   Hint 1: $B\mathbf{x} < \mathbf{0}$ if and only if $B\mathbf{x} + \mathbf{1}z \leq \mathbf{0}, z > 0$.
   Hint 2: Use Farkas' Lemma with a suitable $b \in R^{n+1}$ and $A \in R^{m \times (n+1)}$

3. In this problem, you have to use SSVM code to classify the Adult dataset which is available on UCI Data Repository `www.ics.uci.edu/~mlearn/MLRepository.html` or `http://dmlab8.csie.ntust.edu.tw/#dataset`

   (a) Report the tenfold cross validation training and testing set correctness of linear SSVM. You have to decide the penalty parameter by yourself. (10%)

   (b) Report the tenfold cross validation training and testing set correctness of SSVM with Gaussian kernel. You have to decide the penalty parameter and *width* parameter in Gaussian kernel. (15%)

   (b) Report the tenfold cross validation training and testing set correctness of SSVM with the Reduced Gaussian kernel. You can *randomly* select the reduced set with size 400 and have to decide the penalty parameter and *width* parameter in Gaussian kernel. (15%)

4. Let $A \in R^{10 \times 5}$ be a matrix that represents 10 data points in R^5 input space given as follows:

$$\begin{bmatrix} 1 & 0 & 0 & 2 & -1 \\ 1 & -1 & 2 & 0 & -1 \\ 0 & 1 & 0 & 2 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 2 & 1 & 1 & 2 \\ 0 & -2 & 2 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 2 & 2 & 1 & 1 & 0 \\ 2 & -2 & 1 & 0 & 0 \\ 1 & 1 & 1 & 2 & 1 \end{bmatrix}.$$

Applying *singular value decomposition*(SVD), project these 10 points in R^5 input space onto a 2-dimensional space. Plot these 10 points in the projection 2-dimensional space.

Application of the SVD: Image processing
Suppose a satellite takes a picture, and wants to send it to earth. The picture may contain 1000×1000 *"pixels"*, little squares each with a definite color. We can code the colors, in a range between black and white, and send back $1,000,000$ numbers. It is better to find the essential information in the 1000×1000 matrix, and send *only* that.

Suppose we know the SVD ($A = U \cdot \Sigma \cdot V^T$). The key is in the singular values (in Σ). Typically, some are significant and others are extremely small. If we keep 60 and throw away 940, then we send only the corresponding 60 columns of U and V. The other 940 columns are multiplied in $A = U \cdot \Sigma \cdot V^T$ by the small σ's that are being ignored. In fact, we can do the matrix multiplication as *columns* $\times$ *rows*:

$$U \cdot \Sigma \cdot V^T = u_1 \sigma_1 v_1^T + u_2 \sigma_2 v_2^T + \ldots + u_r \sigma_r v_r^T.$$

If only 60 terms are kept, we send 60×2000 numbers instead of a *million.*
(*Quote from Gilbert Strang's Linear Algebra and its Applications book*)
(30 %)