Homework 2 Answer

Theory [1pt each]

1. Given a sample with feature vector $x = [1.1, 2.2, 3.3]^T$, what is its augmented feature vector?
   - $x = [1.1, 2.2, 3.3, 1]^T$
2. If the weight vector of a linear classifier is $w = [1, 0, 1, 0]^T$, and we define that a sample belongs to class $+1$ if $w^T x > 0$ and -1 if $w^T x < 0$ where x is the augmented feature vector of the sample, what is the class of the sample?
   - $w^T x = 4.4$, so this sample belongs to class $+1$
3. When discussing the sum of error squares loss function in the class, we used augmented but not normalized augmented (normalized and augmented) feature vectors. Please rewrite that loss function $J(\mathbf{W}) = \sum_{i=1}^{N}(\mathbf{x}_i^T\mathbf{W} - y_i)^2$ in terms of **normalized augmented** feature vectors. Let x_i'' be the normalized augmented feature vector of the i-th sample, and w be the weight vector of the classifier. A correct prediction shall satisfy $w^T x_i'' > 0$ regardless of the class of the sample because x_i'' has been normalized. You may use a computational algebra system to help – but it is not required. It might be easier by hand on scratch paper.
   - My thoughts: since the prediction part will always be positive, then the error should always be the difference between two positive numbers. So minus -1 for example in the non-nomarlized function should now be just minus 1. Adding the square to the lable (y_i) can help with this.
   - $J(\mathbf{W}) = \sum_{i=1}^{N}(\mathbf{x''}_i^T\mathbf{W} - y_i^2)^2$
4. Please find the solution for minimizing the new loss function. Keep variables and font style consistent with those in the class notes/slides, except that you can reuse the matrix $\mathbb{X} = \begin{pmatrix} — & \mathbf{x''}_1^T & — \\ — & \mathbf{x''}_2^T & — \\ & \vdots & \\ — & \mathbf{x''}_N^T & — \end{pmatrix}$, each **row** of which is re-purposed into a normalized and augmented feature vector. The right most column of the new $\mathbb{X}$ should contain only 1's and -1's.
   - $\frac{\partial J(\mathbf{W})}{\partial \mathbf{W}} = 2\sum_{i=1}^{N}\mathbf{x}_i(\mathbf{x''}_i^T\mathbf{W} - y_i^2) = (0, \ldots, 0)^T$
   - $\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^T = \mathbb{X}^T\mathbb{X}$
   - $\sum_{i=1}^{N}\mathbf{x}_i y_i^2 = \mathbb{X}^T\mathbf{y^2}$
   - $\mathbb{X}^T\mathbb{X}\mathbf{W} = \mathbb{X}^T\mathbf{y^2}$
   - $(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbb{X}\mathbf{W} = (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbf{y^2}$
   - $\mathbf{W} = (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbf{y^2}$