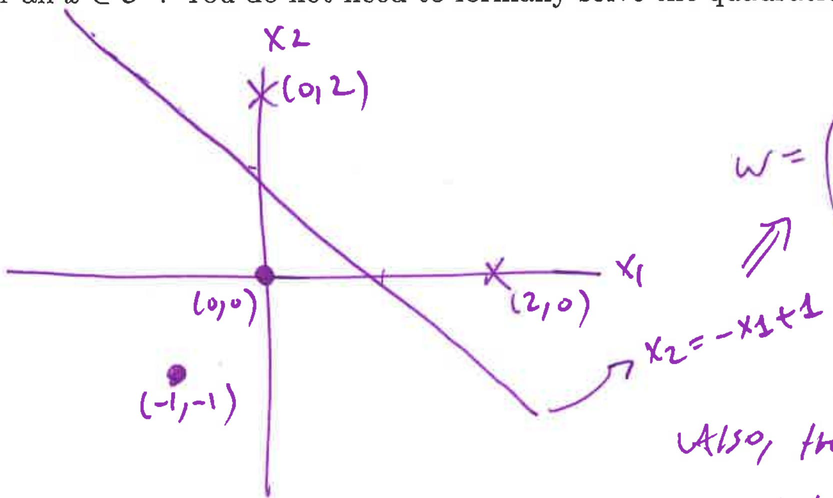
ECE/CS 559 - Fall 2019 - Midterm #2.

Full Name: Luciano Pavarotti **ID Number:** 1935-2007

For all questions, answers without any justification will not be given any credit.

Q1 (15 pts). Consider classes $\mathcal{C}^+ = \{\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}-1\\-1\end{smallmatrix}\right]\}$ and $\mathcal{C}^- = \{\left[\begin{smallmatrix}0\\2\end{smallmatrix}\right], \left[\begin{smallmatrix}2\\0\end{smallmatrix}\right]\}$.

(a) (10 pts) Find w, θ to solve the linear SVM problem with $w^T x + \theta \geq 0$ for all $x \in \mathcal{C}^+$, and $w^T x + \theta < 0$ for all $x \in \mathcal{C}^-$. You do not need to formally solve the quadratic program and can just argue geometrically.



$w = \begin{pmatrix}-1\\-1\end{pmatrix}$ $\theta = +1$

$\Uparrow$

[$w = \begin{pmatrix}1\\1\end{pmatrix}$, $\theta = -1$ seems like another solution but it does not result in the correct outputs]

Also, the above solution is normalized, $w = \begin{pmatrix}-\alpha\\-\alpha\end{pmatrix}$ $\theta = +\alpha$ for any $\alpha > 0$ is also fine.

(b) (5 pts) Indicate the support vectors of the solution in (a).

Support vectors are $\begin{pmatrix}0\\2\end{pmatrix}, \begin{pmatrix}2\\0\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix}$.

Q2 (10 pts). Find the weights and biases of a 2-neuron Hopfield network that satisfies the following properties: (i) The neurons have the sgn($\cdot$) activation function, i.e. $\text{sgn}(x) = 1$, $x \geq 0$, and $\text{sgn}(x) = -1$, $x < 0$. (ii) When the network operates with synchronous updates, every initial state results in a cycle of length 4. For example, if the states are enumerated as u_1, u_2, u_3, u_4, the initial state u_1 would result in the cycle $u_3, u_4, u_2, u_1, u_3, u_4, u_2, u_1, \ldots$, and similarly for the other states.

$W = \begin{pmatrix}0 & 1\\-1 & 0\end{pmatrix}$ $b = \begin{pmatrix}0\\0\end{pmatrix}$ Initial State $\begin{pmatrix}1\\1\end{pmatrix}$ results in the sequence $\begin{pmatrix}1\\1\end{pmatrix} \to \begin{pmatrix}1\\-1\end{pmatrix} \to \begin{pmatrix}-1\\-1\end{pmatrix} \to \begin{pmatrix}-1\\1\end{pmatrix} \to \begin{pmatrix}1\\1\end{pmatrix} \to \cdots$

Q3 (15 pts). Consider n patterns $x_1, \ldots, x_n \in \{-1, 1\}^m$ that we wish to store to an m-neuron autoassociative memory using Hebbian learning and the $\text{sgn}(\cdot)$ activation function, i.e. $\text{sgn}(x) = 1$, $x \geq 0$, and $\text{sgn}(x) = -1$, $x < 0$. Suppose that the patterns are orthogonal in the sense that $x_i^T x_j = 0$ whenever $i \neq j$.

(a) (5 pts) Indicate the weight matrix and the biases of the network as a function of $x_1, \ldots, x_n$.

$$W = \sum_{i=1}^{n} x_i x_i^T.$$

(b) (10 pts) Show that all memory patterns can be recalled perfectly.

$$\text{sgn}(W x_j) = \text{sgn}\left(\sum_{i=1}^{n} x_i x_i^T x_j\right) = \text{sgn}(x_j x_j^T x_j)$$

used orthogonality

$$= \text{sgn}(x_j) = x_j.$$

Q3 (10 pts). Why should we avoid tuning network hyperparameters using the test set?

So that we avoid overfitting to the test set.
"Because testing set should be just for testing purposes" is also fine.

Q4 (10 pts). Explain the dropout method.

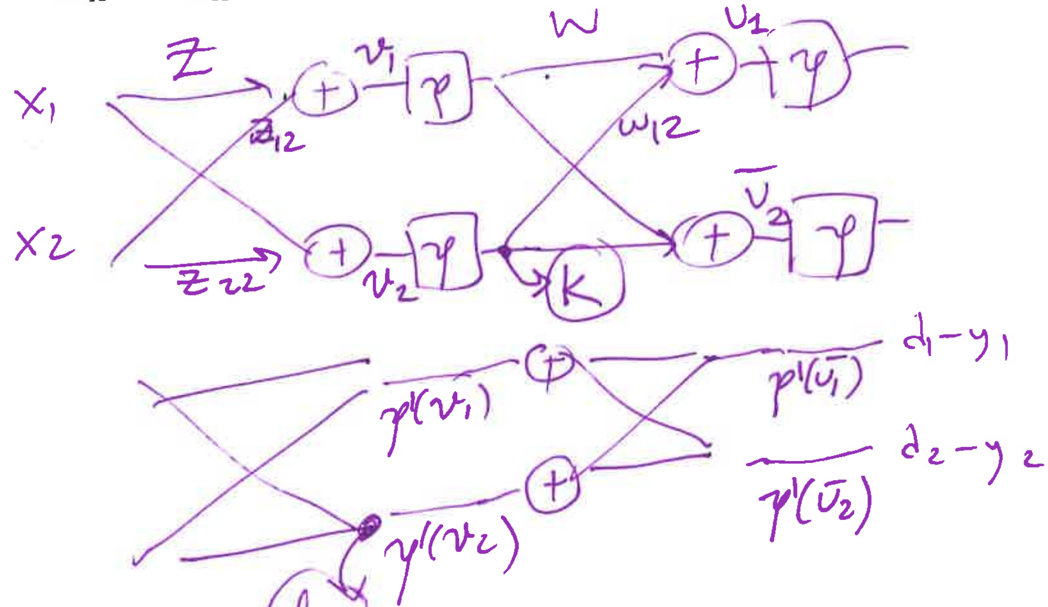
Random removal of network edges/vertices and their restoration during training.

Q5 (10 pts). Consider the minimization $\min_{x \in \mathbb{R}^d} E(x)$, where $E(x)$ is some function and $x = [x_1 \cdots x_d]$. Write down the L_1- and L_2-norm regularized minimizations. Make sure you define all norms explicitly, i.e. you need to write, for example, the L_1-norm of x as a function of the components $x_1, \ldots, x_d$.

$$\mathcal{L}_2 : \min_x \left(\mathcal{E}(x) + \lambda \sum_{i=1}^{d} |x_i|^2 \right)$$

$$\mathcal{L}_2 : \min_x \left(\mathcal{E}(x) + \lambda \sum_{i=1}^{d} |x_i| \right).$$

Q6 (15 pts). You must use the backpropagation algorithm in this question. Let $\mathbf{W} = \left[\begin{smallmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{smallmatrix}\right]$ and $\mathbf{Z} = \left[\begin{smallmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{smallmatrix}\right]$, and $\left[\begin{smallmatrix} y_1 \\ y_2 \end{smallmatrix}\right] = \phi(\mathbf{W}\phi(\mathbf{Z}\left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right]))$, with the understanding that $\phi(\cdot)$ is applied component-wise. Let $E = (d_1 - y_1)^2 + (d_2 - y_2)^2$. Draw the forward and backward propagation graphs and find the partial derivatives $\frac{\partial E}{\partial w_{12}}$ and $\frac{\partial E}{\partial z_{22}}$ in terms of the variables you define on the graphs.



$$\frac{\partial \mathcal{E}}{\partial w_{12}} = -k(d_1 - y_1)\, \varphi'(\bar{v}_1).$$

$$\frac{\partial \mathcal{E}}{\partial z_{22}} = -x_2 l$$

Q7 (15 pts). Consider a two-neuron Hopfield network with weight matrix $W = \begin{bmatrix} +3 & -2 \\ -2 & +1 \end{bmatrix}$, zero biases, and the sgn$(\cdot)$ activation function; i.e., $\text{sgn}(x) = 1$ if $x \geq 0$, and $\text{sgn}(x) = -1$, otherwise.

(a) **(5 pts)** Consider the initial state $x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Find the next state using the synchronous update rule.

$$\text{sgn}\left(\begin{pmatrix} 3 & -2 \\ -2 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right) = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

(b) **(5 pts)** Consider the initial state $x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Does the network state converge with synchronous updates?

$$\text{sgn}\left(\begin{pmatrix} 3 & -2 \\ -2 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix}\right) = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$, so yes it converges to $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

(c) **(5 pts)** Consider the initial state $x = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$. Does the network state converge with asynchronous updates?

Yes. W is symmetric, and has non-negative diagonals.