CPSC 636-600 (Total 100 points)

See the course web page for the **due date** and **submission info**.

Instructor: Yoonsuck Choe

January 31, 2011

1 Learning Rules

Problem 1 (Written: 15 pts): The covariance hypothesis proposed the following weight update based on the pre- ($x_j(n)$) and post-synaptic activity ($y_k(n)$):

$$\Delta w_{kj} = \eta(x_j - \bar{x})(y_k - \bar{y})$$

Show that the expected value of this could be expressed as follows:

$$E[\Delta w_{kj}] = \eta(E[y_k x_j] - \bar{y}\bar{x}).$$

Hint: For a random variable X and its mean μ, the variance is defined as $E[(X - \mu)^2]$. This is equivalent to $E[X^2] - E[X]^2$. The elementary techniques used in the derivation below will help you solve this exercise problem. Note: $E[X] = \mu$, by definition.

$$
\begin{aligned}
E[(X - \mu)^2] &= E[X^2 - 2\mu X + \mu^2] \\
&= E[X^2] - 2\mu E[X] + \mu^2 \\
&= E[X^2] - 2E[X]E[X] + E[X]^2 \\
&= E[X^2] - 2E[X]^2 + E[X]^2 \\
&= E[X^2] - E[X]^2.
\end{aligned}
$$

2 Learning Paradigms

Problem 2 (Written: 5 pts): Given the sigmoid function

$$\phi(v) = \frac{1}{1 + \exp(-av)}$$

show that

$$\phi'(v) = a\phi(v)(1 - \phi(v)).$$

3   Memory

You may use Octave etc. for this part.

Problem 3 (Written: 10 pts):   Given orthonomal key vectors

$$
\begin{aligned}
\mathbf{x}_1 &= [1, 0, 0, 0]^T \\
\mathbf{x}_2 &= [0, 1, 0, 0]^T \\
\mathbf{x}_3 &= [0, 0, 1, 0]^T
\end{aligned}
\tag{1}
$$

The respective stored patterns are

$$
\begin{aligned}
\mathbf{y}_1 &= [10, 3, 1]^T \\
\mathbf{y}_2 &= [-5, 2, 3]^T \\
\mathbf{y}_3 &= [11, 20, 1]^T
\end{aligned}
\tag{2}
$$

(1) Calculate the memory matrix $\mathbf{M}$ and (2) show that perfect mapping is achieved.

Problem 4 (Written: 15 pts):

Consider the memory matrix $\mathbf{M}$ from problem 3 above, and a noisy key

$$\mathbf{x} = [0.2, 0.9, -0.1, 0.15]^T.$$

(1) To which key in problem 3 is $\mathbf{x}$ the closest, in the Euclidean sense? (2) Is the output $\mathbf{y} = \mathbf{M}\mathbf{x}$ similar to the desired output from the clean key identified in (1), again in the Euclidean sense?

Problem 5 (Written: 15 pts):   Consider the following sets of key vectors.

- Set 1:

$$
\begin{aligned}
\mathbf{x}_1 &= [0.78, 0.26, 0.56]^T \\
\mathbf{x}_2 &= [0.71, 0.55, 0.44]^T \\
\mathbf{x}_3 &= [0.59, 0.35, 0.73]^T
\end{aligned}
$$

- Set 2:

$$
\begin{aligned}
\mathbf{x}_1 &= [0.65, 0.76, 0.08]^T \\
\mathbf{x}_2 &= [0.02, 0.04, 1.00]^T \\
\mathbf{x}_3 &= [-0.65, 0.75, 0.09]^T
\end{aligned}
$$

(1) Calculate the pair-wise angle between the vectors in each set: $\mathbf{x}_1$ vs. $\mathbf{x}_2$ in set 1, $\mathbf{x}_1$ vs. $\mathbf{x}_3$ in set 1, $\mathbf{x}_2$ vs. $\mathbf{x}_3$ in set 1. Repeat the calculation for set 2. (2) Based on the results from (1), which set could be more suitable as a key vector set for autoassociative memory? Explain why. (3) Construct the memory matrix for both sets. (4) Plug in the key vectors to the corresponding memory matrix and compare the results (in terms of Euclidean distance between output and desired value).
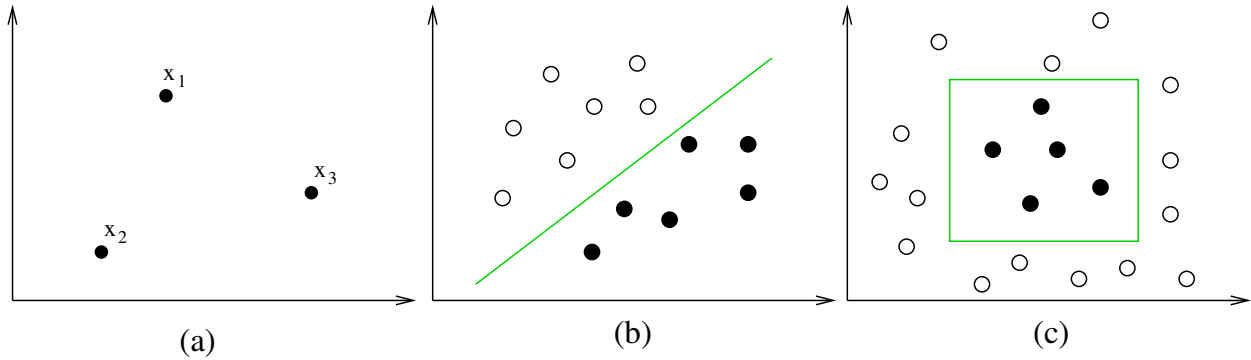
Figure 1: **VC-dimension Example.** Filled and open discs represent inputs of two different classes.

4 Statistical Learning Theory

Problem 6 (Written: 15 pts): (1) List all possible dichotomies for the input set $\{x_1, x_2, x_3\}$ shown above (figure 1a). (2) Can classifiers defined by straight line decision boundaries (see figure 1b) shatter the set in figure 1a? (3) Can classifiers defined by rectangular decision boundaries (assume the rectangles have their sides parallel to the x and the y axis, i.e., you cannot rotate the rectangles: see figure 1c) shatter the set in figure 1a?

Problem 7 (Written: 5 pts): (1) Can the rectangle classifier shatter a set of four inputs? If so, give an example. Otherwise, explain why not. (2) Suppose the rectangle classifier can be rotated. Would a rotatable rectangle classifier have the same VC dimension as the unrotatable one? Or, would it be different. Explain why. (Hint: Can rotatable rectangle classifiers shatter an input set that an unrotatable ones cannot?)

5 Programming exercise

Problem 8 (Written: 5 pts): Show by example that the top and bottom equations in slide02 page 42 are equivalent,

$$\hat{\mathbf{M}} = \sum_{k=1}^{q} \mathbf{y}_k \mathbf{x}_k^T$$

$$\hat{\mathbf{M}} = \begin{bmatrix} \mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_q \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ ... \\ \mathbf{x}_q \end{bmatrix} = \mathbf{Y}\mathbf{X}^T.$$

using the following $\mathbf{X}$ and $\mathbf{Y}$. Follow the example in `src/partmat.m`. You can take out a particular vector from the matrix using the following trick: `k = 3; xk = X(:,k)`. This will give you $[8, 2, 4]^T$.

```
X =

   3   5   8   6   9
   6   6   2   5   3
   5   3   4   7   7

Y =
```

```
   6   9   7   6   7
   4   5   4   4   5
```

Problem 9 (Written: 15 pts):   In Matlab or Octave, follow the instructions below to generate the key matrix X and memory matrix Y. You don't need to understand what it does, at this point.

```
dat = rand(100,3); cv = cov(dat); [X,Xd] = eig(cv); X
Y = rand(5,3)
```

In Matlab or Octave, (1) Calculate the norm of each resulting column vector in X. You can easily do this by X[:,1]'*X[:,1], etc. (2) Check if the column vectors in X are all *nearly* orthogonal with each other. Apply the technique in (1). (3) Construct $\hat{\mathbf{M}}$ as in equation 2.35, and verify that it has the property in equation 2.39 for all the key-vector/memory-vector pairs. Show all your results (print out).