

CPSC 633-600 (Total 100 points)

Machine Learning

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

Instructor: Yoonsuck Choe

February 9, 2009

1 Concept Learning

Problem 1 (Written: 20 pts): Solve exercise 2.4 in the textbook.

2 Perceptron

Problem 2 (Written: 20 pts): Solve exercise 4.2 in the textbook. For each unit (hidden and output), show a graphical representation of the decision boundary.

3 Gradient Descent

Problem 3 (Written: 5 pts): Given an error function

$$E(w) = w^4 - 5w^3 + 5w^2 + 5w - 6,$$

find:

$$\frac{\partial E(w)}{\partial w}.$$

Note that this is simply an ordinary derivative $\frac{dE}{dw}$.

Problem 4 (Written: 5 pts): How many minima does $E(w)$ have? Hint: using $w^4 - 5w^3 + 5w^2 + 5w - 6 = (w + 1)(w - 1)(w - 2)(w - 3)$, you can draw a rough plot from $w = -2$ to 4.

Problem 5 (Written: 5 pts): With $\frac{\partial E(w)}{\partial w}$ calculated above, if you want to adjust w to minimize $E(w)$, what should Δw be? Write the answer as a polynomial function of w . (This is trivial, given the answer to problem 1.)

Problem 6 (Program: 20 pts): Using the gradient found above, write a short program to

1. Initialize w to a particular value w_0 .

2. Repeat

$$w \leftarrow w + \eta \Delta w$$

until change in $E(w) < 0.00001$. Set $\eta = 0.01$.

Experiments: Run your program with $w_0 \in \{-2, 1, 2, 5\}$, and report the following:

1. Plot the function $E(w)$ in the background, and plot the changing $(w, E(w))$ positions as points in the same plot. See figure 1.
2. Report how many iterations were needed for $E(w)$ to reach the minima.

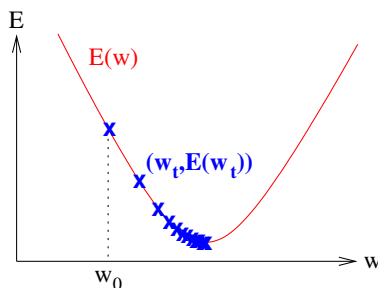


Figure 1: **Gradient Descent.** Note that $E(w)$ here is different from the one in problem 2.

Problem 7 (Program: 10 pts): Add momentum to the previous program and repeat the experiments. Use momentum constant of 0.9. In addition to the required report in the previous program, report if you observe any difference in behavior (which local minimum was reached, etc.).

Problem 8 (Written: 5 pts): Using chain rule, find the gradient for the function $\sigma(E(w))$ where $\sigma(x) = 1/(1 + \exp(-x))$:

$$\frac{d\sigma(E(w))}{dw}.$$

Write the answer as a polynomial function of w . **Hint:** $\sigma'(x) = \sigma(x)(1 - \sigma(x))$.

Problem 9 (Written: 10 pts): Consider these two functions of two variables:

$$E_1(x, y) = x^2 + y^2 \tag{1}$$

$$E_2(x, y) = \frac{x^2}{2} + y^2 \tag{2}$$

1. Illustrate by hand (or use Octave/Matlab) the contour plot of these two functions. Make the x and y range go from -10 to 10, and draw contour lines for $E_i(x, y)$ $i = 1, 2$ values in the increment of 20.
2. Derive the gradient $\nabla E_1(x, y)$ and $\nabla E_2(x, y)$.
3. Pick the contour line at E value (z-axis value) of 40, and draw the gradient vector at 6 different points on that contour (evenly spaced around the contour), based on your derived $\nabla E_1(x, y)$ and $\nabla E_2(x, y)$.
4. Starting from (10, 10), manually perform (you may use a calculator/program) the first four steps of gradient descent (try minimizing E_i while adjusting x and y), with learning rate $\eta = 0.1$, so that

$$\Delta \vec{w} = -\eta \nabla E_i(x, y), \text{ for } i = 1, 2,$$

where $\vec{w} = (x, y)$. Illustrate your results and show the calculated values of $\vec{w}$ over time.

5. Discuss how the behavior differs for the two different functions.