Homework 2

Total possible score: 100 points, 45 points = 100%
Problems 1, 2, and 3 will be graded fully.
Problems 4,5,6,7,8 are for general background refresher. They will receive bonus points:
50% of the respective maximum score if full effort is shown but solution is incorrect;
100% of the score if correct. Bonus points will count toward the overall sum of HW points.

LaTeX users: I posted the .tex source of this homework, to save you some typing.

Problem 1. (13 points total)

1/a (4 points)
If $f(x)$ is the logistic (sigmoid) function

$$f(x) = \frac{1}{1 + e^{-ax}} \tag{1}$$

where a is the slope parameter, show that

$$\frac{df(x)}{dx} = a \cdot f(x)[1 - f(x)]. \tag{2}$$

1/b (4 points)
Calculate the derivative of the hyperbolic tangent function

$$f(z) = \frac{e^{bz} - e^{-bz}}{e^{bz} + e^{-bz}} \tag{3}$$

where b is the slope parameter. Express the derivative as a function of $f(z)$ (similarly as was done for the sigmoid in 1/a) for computational convenience and efficiency.
(The sigmoid and hyperbolic tangent functions are often used assuming $a = 1$ and $b = 1$, respectively. Equations 1 and 2 show more general forms.)

Points 1/a and 1/b show one of the nice properties that makes the sigmoid and the hyperbolic tangent functions suitable for transfer function in an ANN: their derivatives are easy to compute, without performing actual derivation.

I'm the answer

$$\frac{df(z)}{dz} = \frac{b(e^{bz} + e^{-bz})^2 - b(e^{bz} - e^{-bz})^2}{(e^{bz} + e^{-bz})^2} = b(1 - f(z)^2) \tag{4}$$

1/c (5 points)
A Processing Element (PE) has a logistic transfer function with slope parameter a. Assume that $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is the input vector to the PE. For convenience we want to absorb the slope parameter in the inputs so we can use

$$f(x) = \frac{1}{1+e^{-x}} \quad \text{instead of} \quad f(x) = \frac{1}{1+e^{-ax}}. \tag{5}$$

How should we transform the inputs to achieve this?

For Problems 2 and 3 you can write your answers on this assignment page as indicated.

Problem 2. (12 points total)

A PE receives inputs from four other PEs. Let us denote the four PEs sending signal by ${}^1PE_i, i = 1, \ldots, 4$ (PEs #1, 2, 3, and 4 in layer one) and denote the receiving PE by ${}^2PE_j$ (the jth PE in layer two). The activation levels (outputs) ${}^1y_i$ of ${}^1PE_i$ are 10, -20, 4 and -2, respectively, for $i = 1, \ldots, 4$. The weights ${}^2w_{ji}$ of ${}^2PE_j$ (connecting ${}^2PE_j$ to ${}^1PE_i$) are ${}^2w_{j1} = 0.8, {}^2w_{j2} = 0.2, {}^2w_{j3} = -1.0, {}^2w_{j4} = -0.9$. Calculate the output (activation level) ${}^2y_j$ of ${}^2PE_j$ for the cases below. Give your answer as a single number, in-line.

i) (3 points) ${}^2PE_j$ is linear, i.e. the transfer function of the PE is linear. Use f(I) = I.

1.8

ii) (4 points) The transfer function is the hardlimit function.

1

iii) (5 points) The transfer function is the sigmoid with slope parameter $a = 1$.

0.858149

Problem 3. (20 points)

Which of the following functions qualify as a cumulative probability distribution function? Write Yes / No in-line after each formula.

7/a) (5 points)

$$f(x) = \frac{1}{1+e^{-ax}} \quad Yes \quad if \quad a > 0 \tag{6}$$

7/b) (5 points)

$$f(x) = \frac{x}{\sqrt{1+x^2}} \quad No \tag{7}$$

7/c) (5 points)

$$f(x) = \frac{1-e^{-x}}{1+e^{x}} \quad No \tag{8}$$

7/d) (5 points)

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du \quad Yes \tag{9}$$

Problem 4. (15 points total)

In Colin Fyfe's Chapter 2, pages 28–29, it is shown that the distribution with the greatest differential entropy for a given variance is the Gaussian. On page 29 the values of the Lagrange multipliers λ_1 and λ_2 are given without proof. Can you show how to derive these values from the preceding two integral equations on the top of page 29?

Note:
According to my calculation there seems to be a slight error in the formula of λ_1*: it should be equal to* $1 - 0.5 * log(2\pi\sigma^2)$*. You can check this by substituting back to f(x) in (2.11), which should yield the Gaussian.*

I'm the answer

Now, we have

$$\int_{-\infty}^{\infty} exp(-1 + \lambda_1 + \lambda_2(x-\mu)^2)dx = 1 = e^{\lambda_1 - 1} \int_{-\infty}^{\infty} e^{\lambda_2(x-\mu)^2} dx \tag{10}$$

$$\int_{-\infty}^{\infty} (x-\mu)^2 exp(-1 + \lambda_1 + \lambda_2(x-\mu)^2)dx = \sigma^2 = e^{\lambda_1 - 1} \int_{-\infty}^{\infty} (x-\mu)^2 e^{\lambda_2(x-\mu)^2} dx \tag{11}$$

Integrate by parts:

$$\int_{-\infty}^{\infty} (x-\mu)^2 e^{\lambda_2(x-\mu)^2} dx = \int_{-\infty}^{\infty} \frac{1}{2\lambda_2}(x-\mu) e^{\lambda_2(x-\mu)^2} dx - \int_{-\infty}^{\infty} \frac{1}{2\lambda_2} e^{\lambda_2(x-\mu)^2} dx \tag{12}$$

Easy to know the first term of (12) is 0. So (11) could be transformed to

$$-\frac{1}{2\lambda_2} e^{\lambda_1 - 1} \int_{-\infty}^{\infty} e^{\lambda_2(x-\mu)^2} dx = \sigma^2 \tag{13}$$

Then devide (13) by (10), we get:

$$-\frac{1}{2\lambda_2} = \sigma^2 \tag{14}$$

$$\lambda_2 = -\frac{1}{2\sigma^2} \tag{15}$$

Since the Gaussian integral

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi} \tag{16}$$

Insert (15) into (10), we can calculate

$$e^{\lambda_1 - 1} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1 = \sqrt{2}\sigma e^{\lambda_1 - 1} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} d\frac{x-\mu}{\sqrt{2}\sigma} = \sqrt{2}\sigma e^{\lambda_1 - 1}\sqrt{\pi} \tag{17}$$

$$e^{\lambda_1 - 1} = \frac{1}{\sqrt{2\pi}\sigma} \tag{18}$$

$$\lambda_1 = 1 + log(\frac{1}{\sqrt{2\pi}\sigma}) = 1 - 0.5 * log(2\pi\sigma^2) \tag{19}$$

Q.E.D.

Problem 5. (10 points)

Show how to get (2.13) on page 29 of Colin Fyfe's text. That is, calculate the differential entropy of the continuous random variable X whose probability density function is a Gaussian with mean μ and variance σ^2, as in Problem 3.
(Note: In the LHS of definition (2.8) X should stand instead of x. Likewise in (2.13). Further similar inconsistent notations in (2.14). Correct these typos in your copies.)

I'm the answer

$$h(X) = -\int_{-\infty}^{\infty} f(x)logf(x)dx = \int_{-\infty}^{\infty} f(x)(-log(\frac{1}{\sqrt{2\pi}\sigma}) - \frac{(x-\mu)^2}{2\sigma^2})dx$$
$$= -log(\frac{1}{\sqrt{2\pi}\sigma})\int_{-\infty}^{\infty} f(x)dx - \int_{-\infty}^{\infty} \frac{(x-\mu)^2}{2\sigma^2}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$$

Since $\int_{-\infty}^{\infty} f(x)dx = 1$,set $t = \frac{x-\mu}{\sqrt{2}\sigma}$, now we have

$$h(X) = -log(\frac{1}{\sqrt{2\pi}\sigma}) - \int_{-\infty}^{\infty} t^2\frac{1}{\sqrt{\pi}}e^{-t^2}dt$$
$$= -log(\frac{1}{\sqrt{2\pi}\sigma}) - \frac{1}{\sqrt{\pi}}(\int_{-\infty}^{\infty}\frac{1}{2}te^{-t^2}dt - \int_{-\infty}^{\infty}\frac{1}{2}e^{-t^2}dt)$$
$$= -log(\frac{1}{\sqrt{2\pi}\sigma}) + \frac{1}{2}$$
$$= \frac{1}{2}(1 + log(2\pi\sigma^2))$$

Q.E.D.

Problem 6. (10 points total)

5/a (3 points)
Calculate the variance and the differential entropy of the uniform distribution.

I'm the answer

Assume that random variable X is uniformly distributed on [a,b]. Then the pdf of X is :

$$f(x) = \frac{1}{b-a} \quad if \quad x \quad in \quad [a,b]$$
$$= 0 \quad else$$

Then

$$V(X) = E[(X-E(X))^2] = E(X^2) - E(X)^2 = \int_a^b \frac{x^2}{b-a}dx - (\int_a^b \frac{x}{b-a}dx)^2 = \frac{\frac{1}{3}(b^3-a^3)}{b-a} - \frac{\frac{1}{4}(b^2-a^2)^2}{(b-a)^2} = \frac{(b-a)^2}{12} \quad (20)$$

$$h(X) = -\int_a^b \frac{1}{b-a} log(\frac{1}{b-a})dx = log(b-a) \quad (21)$$

5/b (7 points)
Compare the differential entropies of the uniform and Gaussian distributions assuming they have the same variance. Which one is larger? Document your result.

I'm the answer

For unifrom distribution:

$$h(X) = log(b-a) = log(\sqrt{12V(X)}) = \frac{1}{2}(log(12) + log(V(X)) \quad (22)$$

For gausian distribution:

$$h(X) = \frac{1}{2}(1 + log(2\pi\sigma^2)) = \frac{1}{2}(log(e * 2\pi) + log(V(X))) \quad (23)$$

Obviously $e * 2\pi > 12$, so with same variance the Gaussian distribution has larger entropy.

Problem 7. (10 points)

The support of a random variable ξ is $[a, b]$. There is no other constraint. What is the maximum entropy distribution for ξ (what is the probability density function that maximizes the differential entropy of ξ)? Justify your result.

Problem 8. (10 points total)

8/a (5 points) Show that if ξ and η are jointly Gaussian random variables (random variables with joint Gaussian probability density function), and ξ and η are uncorrelated, i.e., $E[\xi\eta] = E[\xi]E[\eta]$ then ξ and η are also statistically independent (a property that is not shared by other probability distributions, in general).

8/b (5 points) Show that if $\mathbf{x} = (x_1, \dots, x_n)$ is a random vector (vector whose elements are random variables), and x_i are statistically independent of each other, then the joint differential entropy of the x_j, $h(\mathbf{x}) = h(x_1, \dots, x_n)$, is:

$$h(\mathbf{x}) = h(x_1) + \cdots + h(x_n)$$

where $h(x_j)$ is the differential entropy of the random variable x_j, for $j = 1, \dots n$.