Intelligent Systems: Reasoning and Recognition

James L. Crowley

MoSIG M1 Final Exam – 11 May 2020

Conditions: You have the right to use any notes, written material or on-line material. You are encouraged to use the documents available on the course web site while completing this exam. You may answer questions in English or in French, but you must illustrate your answers with mathematics and drawings when appropriate. You may answer the questions in any order. Your written answers must be clear and legible. Illegible text will not be graded.

Write your full name on every page and number all pages. Write out and sign the following attestation on the first page by hand: "I, <your full name>, certify that I have not communicated with, or been assisted by, any other person in completing this exam. I acknowledge that violation of this condition would constitute a violation of the academic integrity rules of the UGA and could be subject to penalties, including possible failure or expulsion."

1) (2 points) Provide a definition and an explanation for Precision and Recall. How are they calculated? What do they tell about a classifier?

3) (10 points) You have been hired as a political analyst to work on the political campaign for a referendum. Your job is to identify the sectors of the population for which you can design targeted publicity on Facebook. For this you prepare an on-line personality quiz. Each question has a small number of possible responses. The questions are as follows

1) What is your age? A) 18-29, B) 30-39, C) 40-49, D) 50-59, E) 60 or older
2) How many years of University education do you have? A) None, B) 1 to 2 years, C) 3 to 4 years, D) 5 to 6 years, E) 7 or more.
3) How much do you earn each year? A) < 20K B) 20K to 40K C) >40K to 60K D) >60 to 80K, E) >80K.
4) How will you vote in the referendum? A) Yes, B) No, C) Undecided, D) I will not vote.

a) Explain how to use a ratio of histograms to predict the response to survey question 4 as a function of the answers to survey questions 1, 2 and 3. How large are your histograms? How many people should be polled before you can trust the results. Why?

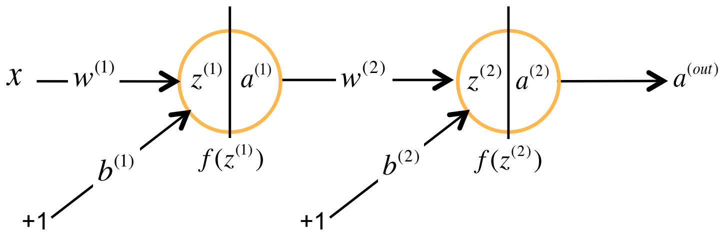
b) Explain how to estimate and use a multivariate normal (Gaussian) density functions to predict the response to question 4 as a function of the answers to survey questions 1, 2 and 3. Be sure to provide the formulas for estimating the mean and covariance, as well as the formulas for estimating the probabilities for each possible response to survey question 4.

c) Explain how to use a Kernel Density Estimator to predict the response to question 4 as a function of the answers to survey questions 1, 2 and 3. Be sure to provide the formulas for estimating the probabilities for each possible response to survey question 4.

d) Explain how to use the K-Nearest neighbors algorithm to predict the response to question 4 as a function of the answers to survey questions 1, 2 and 3. Be sure to provide the formulas for estimating the probabilities for each possible response to survey question 4.

e) How would you calculate the probability of error for any of these estimates?

3) (8 points) The back propagation algorithm is a distributed form of Gradient Descent. To see this, consider the following two-layer network with one neural unit per layer. The network has one input variable, *X*, and one output activation, $a^{(out)}=a^{(2)}$.



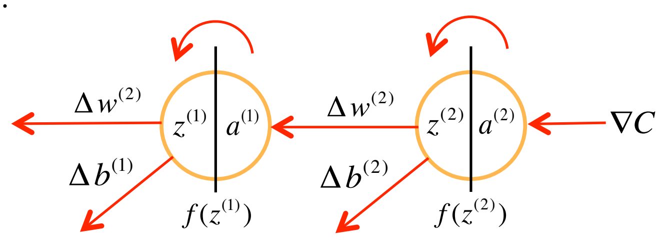
$$\vec{w} = \begin{pmatrix} w^{(1)} \\ b^{(1)} \\ w^{(2)} \\ b^{(2)} \end{pmatrix}$$

This network is described by a vector of four parameters, $\vec{w}$. The error for using these parameters to classify an input sample *X* with indicator variable *y* is

$$\delta^{(out)} = \left(a^{(out)} - y\right).$$

The cost of this error is $C = \frac{1}{2}\left(a^{(out)} - y\right)^2$.

The gradient of the cost, ∇C, tells us how much to adjust each network parameter to correct for this error.



$$\nabla C = \frac{\partial C}{\partial \vec{w}} = \begin{pmatrix} \frac{\partial C}{\partial w^{(1)}} \\ \frac{\partial C}{\partial b^{(1)}} \\ \frac{\partial C}{\partial w^{(2)}} \\ \frac{\partial C}{\partial b^{(2)}} \end{pmatrix} = \begin{pmatrix} \Delta w^{(1)} \\ \Delta b^{(1)} \\ \Delta w^{(2)} \\ \Delta b^{(2)} \end{pmatrix}$$

To evaluate these derivatives, we use the chain rule from calculus. For example, the correction terms for $w^{(2)}$ and $b^{(2)}$ are

$$\Delta w^{(2)} = \frac{\partial C}{\partial w^{(2)}} = \frac{\partial C}{\partial a^{(2)}} \cdot \frac{\partial a^{(2)}}{\partial z^{(2)}} \cdot \frac{\partial z^{(2)}}{\partial w^{(2)}} \quad \text{and} \quad \Delta b^{(2)} = \frac{\partial C}{\partial b^{(2)}} = \frac{\partial C}{\partial a^{(2)}} \cdot \frac{\partial a^{(2)}}{\partial z^{(2)}} \cdot \frac{\partial z^{(2)}}{\partial b^{(2)}}$$

The correction terms for $w^{(1)}$ and $b^{(1)}$ are

$$\Delta w^{(1)} = \frac{\partial C}{\partial w^{(1)}} = \frac{\partial C}{\partial a^{(2)}} \cdot \frac{\partial a^{(2)}}{\partial z^{(2)}} \cdot \frac{\partial z^{(2)}}{\partial a^{(1)}} \cdot \frac{\partial a^{(1)}}{\partial z^{(1)}} \cdot \frac{\partial z^{(1)}}{\partial w^{(1)}} \quad \text{and} \quad \Delta b^{(1)} = \frac{\partial C}{\partial b^{(1)}} = \frac{\partial C}{\partial a^{(2)}} \cdot \frac{\partial a^{(2)}}{\partial z^{(2)}} \cdot \frac{\partial z^{(2)}}{\partial a^{(1)}} \cdot \frac{\partial a^{(1)}}{\partial z} \cdot \frac{\partial z^{(1)}}{\partial b^{(1)}}$$

We can simplify these expressions by computing an error term, δ, for each neural unit. These error terms can be computed recursively, working back from $\delta^{(out)}$ to derive $\delta^{(2)}$ for unit 2 as

$$\delta^{(2)} = \frac{\partial C}{\partial a^{(2)}} \cdot \frac{\partial a^{(2)}}{\partial z^{(2)}} = \delta^{(out)} \frac{\partial f(z^{(2)})}{\partial z^{(2)}}$$

and to derive $\delta^{(1)}$ for unit 1 as

$$\delta^{(1)} = \frac{\partial C}{\partial a^{(2)}} \cdot \frac{\partial a^{(2)}}{\partial z^{(2)}} \cdot \frac{\partial z^{(2)}}{\partial a^{(1)}} \cdot \frac{\partial a^{(1)}}{\partial z^{(1)}} = \delta^{(2)} \cdot w^{(2)} \cdot \frac{\partial f(z^{(1)})}{\partial z^{(1)}}$$

This can be generalized for any number of layers, L, with formulas for the output level L as:

$$\delta^{(L)} = \frac{\partial C}{\partial a^{(L)}} \cdot \frac{\partial a^{(L)}}{\partial z^{(L)}} = \delta^{(out)} \frac{\partial f(z^{(L)})}{\partial z^{(L)}}$$

And for all other units for levels *l=1* to *L-1*

$$\delta^{(l)} = \frac{\partial C}{\partial a^{(L)}} \cdot \frac{\partial a^{(L)}}{\partial z^{(L)}} \cdot \frac{\partial z^{(L)}}{\partial a^{(L-1)}} \cdot \ldots \cdot \frac{\partial z^{(l+1)}}{\partial a^{(l)}} \cdot \frac{\partial a^{(l)}}{\partial z^{(l)}} = \delta^{(l+1)} \cdot w^{(l+1)} \cdot \frac{\partial f(z^{(l+1)})}{\partial z^{(l+1)}}$$

To show this for the simple 2-layer network, answer the following questions:

a) Write the equations for $z^{(1)}$, $a^{(1)}$, $z^{(2)}$, $a^{(2)}$.

b) Show that $\delta^{(out)} = \frac{\partial C}{\partial a^{(2)}}$

c) Given that $\delta^{(2)}$ is defined as $\delta^{(2)} = \frac{\partial C}{\partial a^{(2)}} \cdot \frac{\partial a^{(2)}}{\partial z^{(2)}}$, show that $\delta^{(2)} = \delta^{(out)} \frac{\partial f(z^{(2)})}{\partial z^{(2)}}$

d) Show that $\Delta w^{(2)} = \delta^{(2)} \cdot a^{(1)}$

e) Show that $\Delta b^{(2)} = \delta^{(2)}$

f) Given that $\delta^{(1)}$ is defined as $\delta^{(1)} = \frac{\partial C}{\partial a^{(2)}} \cdot \frac{\partial a^{(2)}}{\partial z^{(2)}} \cdot \frac{\partial z^{(2)}}{\partial a^{(1)}} \cdot \frac{\partial a^{(1)}}{\partial z^{(1)}}$ show that $\delta^{(1)} = \delta^{(2)} \cdot w^{(2)} \cdot \frac{\partial f(z^{(1)})}{\partial z^{(1)}}$

g) Show that $\Delta w^{(1)} = \delta^{(1)} \cdot x$

h) Show that $\Delta b^{(1)} = \delta^{(1)}$