Pattern Recognition and Machine Learning

James L. Crowley

ENSIMAG 3 Final Exam - 8 Feb 2021

Règles et Conditions

Il est interdit de communiquer avec tout personne autre que le Professeur Crowley entre le moment que vous commencer cette examen et le moment que vous rendez vous solutions par courrier électronique.

Vous avez le droit d'utiliser toutes notes, documents écrits ou documents trouvés en ligne, mais il faut citer toutes vos sources. Vous êtes encouragé à utiliser les documents disponibles sur le site Web du cours. Vous pouvez répondre aux questions en anglais ou en français, mais vous devez illustrer vos réponses avec des mathématiques et des dessins, le cas échéant.

Écrivez votre nom complet sur chaque page et numérotez toutes les pages. Votre examen terminé doit être retourné sous forme de fichier .pdf envoyé par email à James.Crowley@grenoble-inp.fr.

Vous pouvez utiliser un logiciel d'édition tel que LaTex ou MS Word, mais sachez que la plupart des questions nécessitent l'écriture de mathématiques. Vous pouvez également écrire vos réponses sur papier et envoyer une copie numérisée ou photographiée au format .pdf. Les copies numériques, ainsi que vos réponses écrites, doivent être claires et lisibles.

Rédigez et signez l'attestation suivante à la fin de votre examen:

Je, <votre nom complet>, certifie que je n'ai pas communiqué avec une autre personne ni été aidé par une autre personne pour compléter cet examen. Je reconnais que toutes infraction à cette condition constituerait une violation des règles d'intégrité académique de Grenoble INP et pourrait être passible de sanctions, y compris d'éventuels échec de l'examen ou expulsion.

Rules and Instructions

You may not communicate with any person by any means while completing this exam.

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. You may answer questions in English or in French, but you must illustrate your answers with mathematics and drawings when appropriate.

Write your full name on every page and number all pages. Your completed exam should be submitted as a .pdf file sent by email to James.Crowley@grenoble-inp.fr

You may use a document typesetting program such as LaTex or MS Word, but beware that most questions require writing mathematics. Alternatively you may write out your answers on paper and send a scanned or photographed copy as a .pdf. Your written answers must be clear and legible.

Write out and sign the following attestation at the end of your exam: "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 Grenoble INP and could be subject to penalties, including possible failure or expulsion."

1) (4 points) While training a neural network, after each epoch of training with a set of training data, you evaluate the resulting network with a separate set of evaluation data. After a certain number of epochs, you observe that the loss for the training set continues to improve while the loss for the evaluation data grows larger. What does this tell you about your network?

2) (4 points) You are provided with a Viola-Jones style face detector composed of a committee of weak classifiers trained with Ada boost. The committee has been trained to accept 10% false positives and 1% false negatives. Is it possible to determine the accuracy for a detector that meets this criteria from the ROC score? If yes, give the formula. If no, explain why.

3) (4 Points) You are asked to use a linear Support Vector machine to construct a pattern detector using 3 dimensional feature vectors $\vec{X}$. You have used an on-line SVM learning algorithm that uses Lagrange Multipliers to discover the support vectors $\{\vec{X}_s\}$.

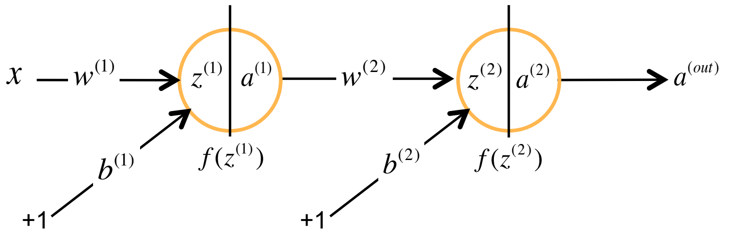
a) How many support vectors should the learning algorithm provide?

b) How are the coefficients $\vec{W}$ and b computed from the support vectors $\{\vec{X}_s\}$?

c) Write the equation for the discriminant function $g(\vec{X})$.

d) What happens if the data are not separable?

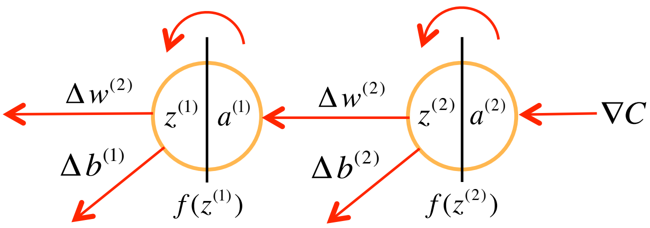
4) (8 points) Back propagation is a distributed form of Gradient Descent. To see this, consider the following two-layer network with one neural unit per layer. The network has one input variable, *X*, and one output activation, $a^{(out)}=a^{(2)}$.



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 (or Loss) for this error is $C = \frac{1}{2}\left(a^{(out)} - y\right)^2$. The gradient of the cost, ∇C, tells how much to adjust each network parameter to correct for the 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 $\Delta b^{(1)} = \delta^{(1)}$