Artificial Intelligence - 2017
Machine Learning, Connectionism
and Data Analytics

Prepared By;
Vijani Supeshala Piyawardana | BSc (Hons) Computer Science |
Demonstrator - COMP3004L, COMP3008L - UCD |
NSBM Green University Town

Important topics;

- empirical error of the hypothesis
- consistent hypotheses for training samples
- regularisation term in a machine learning
- Supervised machine learning - Loss function
- Unsupervised machine learning
- Classification
- Training set, test set

Machine Learning, Connectionism and Data Analytics

Supervised Machine Learning
Unsupervised Machine Learning
Neural Networks
Bayesian Methods

What is Machine Learning

The term Machine Learning refers to “a scientific discipline that explores the construction and study of algorithms that can learn from data”

In general, there are two broad ways that we can learn from data:

1. Supervised learning
2. Unsupervised learning

Machine Learning Problems

Three machine learning problems / three common types of machine learning problems

For supervised learning

1. Classification
2. Regression

For unsupervised learning

3. Clustering

Classification Problem

Goal : predict category of new observation
Based on earlier observations of how the input maps to the output, classification tries to estimate a classifier that can generate an output to arbitrary input.

Earlier Observations ----------Estimate----------> CLASSIFIER

A classifier will classify unseen data to a class

Unseen Data ----------CLASSIFIER----------> Class

Applications for classification

Medical Diagnosis : classes SICK / NOT SICK
Animal Recognition : classes DOG / CAT / HORSE
Spam mail detection problem : SPAM / NOT SPAM

The important thing here is :

The output is qualitative
Predefined classes

Regression Problem

Kind of machine learning problem that tries to predict a continuous or a qualitative value to an input based on previous information. Input variables are called the predictors and the output is the response.

PREDICTORS ----------REGRESSION FUNCTION ----------> RESPONSE

Regression is pretty similar to classification. You are trying to estimate a function that map input to output based on earlier observations. But here you are trying to estimate an actual value, not just a class on observation.

Example: We have a data set on a group of people's height and weight. The question is, Is there a relationship between the height and the weight?
Will a change in height correlate linearly with a change in weight?
If so, can you predict a height of a new person given their wight?

These questions can be answered with linear regression.
Fit a linear function, between the PREDICTOR (the weight) and the RESPONSE (the height)

Regression applications:

- Modelling credit scores based on past payments
- Find the trend of your youtube subscriptions over time
  (TIME ---> SUBSCRIPTIONS)
- Estimate your chances on landing a job at your favourite company based on your college grades. (GRADES ---> LANDING A JOB)

The response or the thing you're trying to predict is always quantitative.
You'll always need input knowledge on previous input to output observations in order to build your model.

Clustering problem

You are trying to group objects that are similar in clusters.
Similar within a cluster.
Dissimilar between clusters.

Supervised Learning

- Learning with a teacher.
- Learn by examples, inputs with labels
- Teacher provide the neural network with desired response for that training input.
- The network parameters are adjusted under the combined influences of training input and error signal.
- The error signal is defined as the difference between the actual response and desired response of the network
- This adjustment is carried out iteratively step by step fashion with the aim of eventually making the network emulate the teacher

Classification and regression are quite similar, for both you try to find a function f or a model which can later be used to predict a class/label or a value to unseen observations.
During the training of the function, labeled observations are available/ are given to the algorithm. These techniques are called, supervised learning.

Supervised Learning

- Supervised methods learn by generalising from training data in the form of sets of inputs with desired outputs.
- learning to predict output when examples of input and corresponding output are available.

A set of examples: <x,f(x)>

x is some object (instance) $\in$ X (instance space)

$f(\cdot)$ is an unknown function

Learning algorithm will guess $h(\cdot) \approx f(\cdot)$

Inductive learning hypothesis

Any $h(\cdot)$ that approximates $f(\cdot)$ well on training examples will also approximate $f(\cdot)$ well on new (unseen) instances x.

f(x) is a real value/an array of real values : regression

f(x) is a discrete value; $f(x) \in \{y\ 1\ ..y\ K\}$: classification

if K=2, binary classification

Example

- **Medical diagnosis (classification)**
  - **x: set of properties for a patient (symptoms, lab tests, previous diseases)**
  - **f(x): disease**
  - **<x,f(x)>: Database of past medical records**
  - **Type of x: a fixed-length array (attribute/value representation), maybe with missing attributes**
  - **Type of f(x): a discrete value or a fixed-width array**

- **Training examples: <x,f(x)>**
- **Hypothesis space: set H of functions that (hopefully) contains the target function f(·)**
- **Result of learning: a function h(·)∈H approximating f(·)**
- **Examples are used to guide the algorithm to choose a good h(·)∈H**

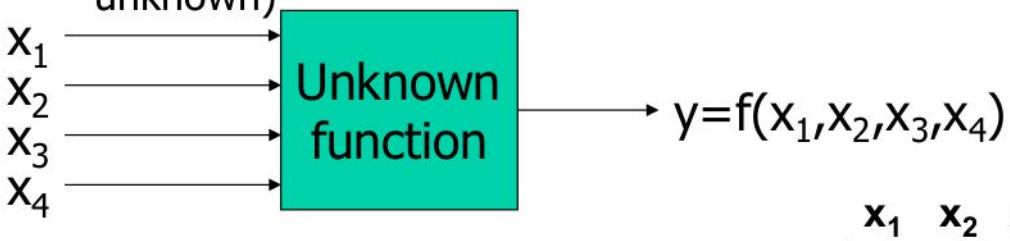
- **h(·) is called a *consistent hypothesis* if it agrees with f(·) on all training examples**
  - **After observing the data, only some hypotheses in H are consistent. They form the so called version space : {h(·)∈H : h(x)=f(x) ∀ example x}**
- **A *consistent learner* always outputs a consistent hypothesis (i.e., it is 100% accurate on the training set)**
- **The empirical error is the fraction of training examples such that h(x)≠f(x) (0% in a consistent learner)**

Example : Boolean Function

Four boolean variables as input features:

$x=(x_1,x_2,x_3,x_4)\in\{0,1\}^4$

$f(x_1,x_2,x_3,x_4) = \neg x_2 \wedge x_4$ (but it is unknown)



Training examples

x_1	x_2	x_3	x_4	y
0	0	1	1	1
1	0	0	1	1
1	0	1	1	1
1	1	0	0	0
1	1	1	0	0

Hypothesis space

- **No knowledge: H is the set of all boolean functions on 4 variables. $|H| = 2^{16} \approx 64 \times 10^3$**

x_1	x_2	x_3	x_4	Y
0	0	0	0	?
0	0	0	1	?
0	0	1	0	?
0	0	1	1	?
0	1	0	0	?
0	1	0	1	?
0	1	1	0	?
0	1	1	1	?
1	0	0	0	?
1	0	0	1	?
1	0	1	0	?
1	0	1	1	?
1	1	0	0	?
1	1	0	1	?
1	1	1	0	?
1	1	1	1	?

- After observing 5 examples we are left with $2^{16-5}=2^{11}=2048$ consistent hypotheses
- A randomly-picked consistent h(·) is unlikely to approximate f(·) well…
- What is wrong?

x_1	x_2	x_3	x_4	Y
0	0	0	0	?
0	0	0	1	?
0	0	1	0	?
0	0	1	1	1
0	1	0	0	?
0	1	0	1	?
0	1	1	0	?
0	1	1	1	?
1	0	0	0	?
1	0	0	1	1
1	0	1	0	?
1	0	1	1	1
1	1	0	0	0
1	1	0	1	?
1	1	1	0	0
1	1	1	1	?

- A literal is a variable x_i or its negation $\neg x_i$
- A term is the conjunction of literals
- H= {terms over x_1,x_2,x_3,x_4}
- $|H|=3^4=81$ (each variable is affirmed, is negated, isn't present)
- Learning algorithm:
  - Initial h(·) = conjunction of all possible literals
  - Remove literals associated with inconsistent *positive* examples

Learning Conjunctions

1. $h = \neg x_1 x_1 \neg x_2 x_2 \neg x_3 x_3 \neg x_4 x_4$
2. Observe 1st training example, remove literals x_1, x_2, $\neg x_3$, and $\neg x_4$
   $h = \neg x_1 \neg x_2 x_3 x_4$
3. Observe 2nd training example, remove literals $\neg x_1$ and x_3
   $h = \neg x_2 x_4$
4. Observe 3rd training example: nothing to do ($h = \neg x_2 x_4$)
5. No more positive training examples
   Output $h = \neg x_2 x_4$

x_1	x_2	x_3	x_4	y
0	0	1	1	1
1	0	0	1	1
1	0	1	1	1

Learning as refinement

- **Start with a small hypothesis class (e.g., boolean conjunctions)**
  - **this means we need to *know a priori* something about the solution**
- **Use examples to infer the particular function (e.g. conjunction) in the class.**

Unsupervised Learning

- Learning without a teacher
- Learn by data, inputs without labels
- No teacher or critic to oversee the learning process
- Weight adjustments are carried out iteratively step by step fashion with the aim of eventually making the network/ machine identify the statistical properties in the data set (ex: mean, spread)

Labeling can be a tedious work and it is often done by human. There are other techniques which don't require labeled data/observations.
These techniques are called unsupervised learning.
.
Unsupervised Learning

- Unsupervised methods learn from input data, which has no desired outputs associated with it. Instead these method search for structure in the input data.
- I want to create an internal representation of the data, e.g. in form of clusters, extract relevant features for further tasks, etc.

Clustering is an unsupervised learning technique.
Clustering : find group of observations that are similar and does not require specific labeled observations.

Training Data
A **training** set is a set of **data** used to discover potentially predictive relationships.
A test set is a set of **data** used to assess the strength and utility of a predictive relationship.
Test and **training** sets are used in intelligent systems, machine learning, genetic programming and statistics.

Reinforcement Learning

- No teacher to provide a desired response at each step of the learning process.
- The learning of input output mapping is performed through continued interaction with the environment in order to minimize a scalar index of performances
- The learning machine must be able to assign credit and blame individually to each action in the sequence of time steps that is led to the final outcome, while the primary reinforcement may only evaluate the outcome.

- Another important branch of machine learning is reinforcement learning, in which a program must interact with a dynamic environment to perform a certain goal or task.
- no real supervision, only occasional payoff: e.g. I don't know if a chess move is correct but I know if I win.

Past exam questions

2014 AI Exam

Question 5

(a) In the context of supervised machine learning, what is an hypothesis and what is the hypothesis space? What is a consistent learner? (6 marks)

(b) What is the empirical error of the hypothesis, $!x_2 \wedge x_4$ when learning a boolean function on the training samples given below: (6 marks)

x_1	x_2	x_3	x_4	y
0	0	1	1	1
1	0	0	1	1
1	0	1	1	1
1	1	1	1	1
1	1	1	0	0

The empirical error is the count of the number of training instances for which the hypothesis gives an incorrect result. i.e. the number of training instance for which (!x_2 ^ x_4) != y

The only training example for which the hypothesis and y do not agree is the 4th one (out of 5)

So the empirical error is 1/5.

(c) Explain the following terms:
- (i) Feature vector
- (ii) Loss function
- (iii) Overfitting
- (iv) Regularisation (8 marks)

2015 AI Exam

Question 5

(a) Distinguish between supervised and unsupervised machine learning. (6 marks)

(b) In the context of supervised machine learning, what is an hypothesis and what is the hypothesis space? What is a consistent learner? (6 marks)

(c) Explain the following terms:

(i) Feature vector

(ii) Loss function

(iii) Overfitting

(iv) Regularisation (8 marks)

2016 AI Exam

Question 5 (a) Clearly define a *classification* problem in the context of Machine Learning.

(2 marks)

(b) Distinguish between *supervised* and *unsupervised* machine learning.

(4 marks)

(c) In the context of supervised machine learning, what is an hypothesis and what is the hypothesis space? What is a consistent learner?

(6 marks)

(d) Explain the following terms:

i. Feature vector

ii. Loss function

iii. Overfitting

iv. Regularisation

(8 marks)

Answer

Question 5 (a) A classification problem : given an input data, determine what category or class the input belongs e.g. input could be an email and the categories could be spam and not-spam

(2 marks)

(b) Supervised = methods that generalise from training data, in the form of inputs with desired outputs

Unsupervised = methods that learn from input data that has no desired output associated with it.
(2 marks for each definitions – main distinction being the desired/ no desired output)

(4 marks)

(c) An hypothesis is a function that approximates the function that we want to learn (2 marks). The hypothesis space is the set of functions from which the hypothesis is drawn (2 marks). A consistent learner is one that learns a hypothesis that matches the true function on all training instances (2 marks)

(6 marks)

(d) Explain the following terms: (2 marks each)

i. Feature vector = representation of an instance as an n-dimensional set of features, where n is the number of features used to represent instances in the learning problem.

ii. Loss function = a function that represents the loss or penalty for getting a prediction wrong.

iii. Overfitting = tendency of a learning algorithm to find regularities or patterns in the noise of the training data, rather than the true patterns.

iv. Regularisation = method to avoid overfitting by imposing some structure on the set of allowed solutions.

(8 marks)

2017 exam

Question 5 (a) Choose the correct answer from the following multiple choice questions: **(10 marks)**

i. What is the empirical error of the hypothesis, $\neg x_2 \vee x_4$ when learning a boolean function on the training samples given below?

x_1	x_2	x_3	x_4	y
0	0	1	1	1
1	0	0	1	1
1	0	1	1	1
1	1	1	1	1
1	1	1	0	0

A. 1
B. 2
C. 3
D. 4

ii. Which of the following hypotheses is *consistent* for these training samples?

A. x_4
B. x_1
C. $\neg x_2$
D. $\neg x_2 \vee \neg x_4$

iii. A regularisation term in a machine learning optimisation objective is included in order
   A. To exploit regularities in the feature data.
   B. To avoid irregularities in the feature data.
   C. To obtain a solution that performs well on training data.
   D. To obtain a solution that performs well on test data.

iv. In the context of supervised machine learning, a *loss function* is
   A. a function that counts the number of training instances for which an incorrect prediction is made
   B. is a measure of the overall performance of the learning algorithm
   C. a function that calculates the penalty associated with making an incorrect prediction
   D. is a measure of the amount of training data that is available to the learning algorithm

v. Which of the following loss functions penalises a mistake in the classification most severely:
   A. Hinge Loss
   B. Square Loss
   C. It depends on the value of the error.
   D. Logistic Loss

(b) Using spam mail detection as an example, explain how machine learning can be applied to solve a classification problem. Discuss how a feature vector would be formed for the spam mail problem.

(6 marks)

(c) In machine learning, what are the training and test sets? Describe two ways that the performance of a classifier can be evaluated using the test set.

(4 marks)

Vocabulary

Discipline - විනය, නීතිය, පුහුණුව
Hypothesis - උපකල්පිතය a proposed explanation made on the basis of limited evidence as a starting point for further investigation.

What is a Neural Network?

- Biological inspiration
- Computational equivalent

In human brain each neuron receives thousands of connections with other neurons, constantly receiving incoming signals to each cell body.

If the resulting sum of the signals surpasses a certain threshold, a response is sent through the axon.
An artificial neural network attempts to recreate the computational mirror of the biological neural network.
Artificial neurons, nodes, connected to each other, the strength of their connections to one another is assigned a value based on their strength , inhibition - maximum being -1.0 or excitation - maximum being +1.0.
If the value is high, the connection is strong.
Within each node, a transfer function is built in. three types of nodes, input, hidden and output.
Based on the connection strength (weights), inhibition or excitation and transfer functions, the activation value is passed from node to node.
Each node sums the activation values it receives, it then modifies the value based on its transfer function.
The activation flows through the network, through hidden layers until it reaches the output nodes.
The output nodes then reflect the input in a meaningful way to the outside world.