Greetings, these are my notes for week 2 of the PadhAI course, covering the topics of MP Neuron and Perceptron.

# 

[**2.1: MP Neuron**](#_ao91ha8bwkbh) **1**

[2.1.1: MP Neuron Model](#_h0ljldv4l75o) 1

[2.1.2: MP Neuron Data and Task](#_fb8nsikrsos0) 2

[2.1.3: MP Neuron Loss](#_ql3nazxnrhuv) 3

[2.1.4: MP Neuron Learning Algorithm](#_v3fdjmoac8jm) 4

[2.1.5: MP Neuron Evaluation](#_eyk6ffudum5q) 5

[2.1.6: Geometric Basics](#_ma4elr2or6on) 6

[2.1.7: MP Neuron Geometric Interpretation](#_c50emz6m22dw) 6

[2.1.8: Summary](#_jn464qn8pu62) 6

[**2.2: Perceptron**](#_rai03ml7gcp5) **7**

[2.2.1: Perceptron Data and Task](#_rtprtpecawaf) 7

[2.2.2: Perceptron Model](#_5v930ae0pnmf) 8

[2.2.3: Perceptron Geometric Interpretation](#_ymwgdmxq6fbk) 9

[2.2.4: Perceptron Loss Function](#_2tklb5r0t99) 9

[2.2.5: Perceptron Learning - General Recipe](#_3rzlvzdkmuoj) 10

[2.2.6: Perceptron Learning Algorithm](#_hhfiezmpxkqr) 10

[2.2.7: Perceptron Learning - Why it works?](#_jkocrz8311bf) 11

[2.2.8: Perceptron Learning - Will it always work](#_3l7m8hpwo486) 11

[2.2.9: Perceptron Evaluation](#_f7e2b594zlux) 12

[2.2.10: Perceptron Summary](#_6kfwn0mofmlu) 13

[2.1.11: Perceptron: Toy Example](#_q1gwqww16ewx) 13

# 2.1: MP Neuron

Artificial Neuron:

* Takes in a series of inputs[x0..xn], adds weights [θ0..θn], applies a function like the sigmoid function and returns the transformed output

McCulloch-Pitts Neuron:

* The early model on an artificial neuron is introduced by Warren McCulloch(neuroscientist) and Walter Pitt(logician) in 1943
* The McCulloch-Pitts Neuron is also known as a linear threshold gate

We are going to be applying the 6-jars framework to learn about the MP neuron

## 2.1.1: MP Neuron Model

What is the mathematical model?



1. Inputs belong to the discrete set of values {0,1}
2. g aggregates the inputs and function f takes a decision based on these aggregations
3. These inputs can be excitatory or inhibitory
   1. y = 0 if xi is inhibitory (outputs zero, sort of an override), else
   2. g(x) = ni=1xi
   3. y = f(g(x))
      1. y = 1 if g(x) >= b
      2. y = 0 if g(x) < b
      3. Where b is a threshold value
      4. b is a parameter, it is adjusted with the aim of maximizing the number of correct predictions

## 2.1.2: MP Neuron Data and Task

What kind of data and tasks can MP neuron process

1. Consider the example of detecting whether the cricketer is out by LBW

|  |  |  |  |
| --- | --- | --- | --- |
| Pitch in line(x1) | Impact(x2) | Missing Stumps(x3) | Is it LBW(y) |
| 1 | 0 | 0 | 0 |
| 0 | 1 | 1 | 0 |
| 1 | 1 | 1 | 1 |
| 0 | 1 | 0 | 0 |

1. We are interested in finding out the relationship between y and xi
2. Here we use y = (3i=1xi >= b)
   1. y = 1 if g(x) >= b
   2. y = 0 if g(x) < b
3. In case our data has non-boolean inputs, we can convert them to a boolean form
4. For example, consider the following boolean-ised phone spec data

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | phone1 | phone2 | phone3 | phone4 | phone5 | phone6 | phone7 | phone8 | phone9 | phone 10 |
| Launch (within 6 months) x1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 |
| Weight (<160g) x2 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| Screen Size (< 5.9in) x3 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| Dual sim x4 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| Internal mem(>= 64gb, 4gb ram) x5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| NFC x6 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 |
| Radio x7 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| Battery (>= 3500mAh) x8 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 |
| Price? (> 20k) x9 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| Liked (y) | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 |

## 2.1.3: MP Neuron Loss

How do we compute the loss

1. Consider the previous example

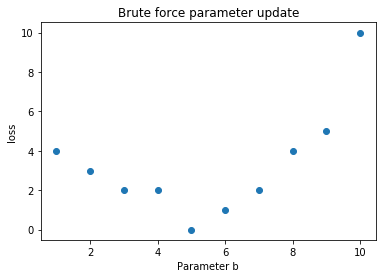
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | phone1 | phone2 | phone3 | phone4 | phone5 | phone6 | phone7 | phone8 | phone9 | phone 10 |
| Launch (within 6 months) x1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 |
| Weight (<160g) x2 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| Screen Size (< 5.9in) x3 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| Dual sim x4 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| Internal mem(>= 64gb, 4gb ram) x5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| NFC x6 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 |
| Radio x7 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| Battery (>= 3500mAh) x8 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 |
| Price? (> 20k) x9 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| Liked (y) | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 |
| **Prediction** | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| **loss** | 0 | 0 | 1 | -1 | 0 | 0 | -1 | 1 | 0 | 0 |
| **Square error** | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 |

1. Take the square of the difference to ignore the sign.
2. cost/loss = i(yi - **i)**2

## 2.1.4: MP Neuron Learning Algorithm

How do we train our model

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | phone1 | phone2 | phone3 | phone4 | phone5 | phone6 | phone7 | phone8 | phone9 | phone 10 |
| Launch (within 6 months) x1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 |
| Weight (<160g) x2 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| Screen Size (< 5.9in) x3 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| Dual sim x4 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| Internal mem(>= 64gb, 4gb ram) x5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| NFC x6 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 |
| Radio x7 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| Battery (>= 3500mAh) x8 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 |
| Price? (> 20k) x9 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| Liked (y) | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 |
| **Prediction** | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? |

1. = (ni=1xi >= b)
2. cost/loss = i(yi - **i)**2
3. In this case, we have only one parameter, so we can afford to use brute force search.
   1. Here, consider we have n features
   2. b can only range from 0 to n, else it would be a pointless parameter
   3. b has discrete values only, as the inputs are also discrete values

## 2.1.5: MP Neuron Evaluation

1. Training Data

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | phone1 | phone2 | phone3 | phone4 | phone5 | phone6 | phone7 | phone8 | phone9 | phone 10 |
| Launch (within 6 months) x1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 |
| Weight (<160g) x2 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| Screen Size (< 5.9in) x3 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| Dual sim x4 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| Internal mem(>= 64gb, 4gb ram) x5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| NFC x6 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 |
| Radio x7 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| Battery (>= 3500mAh) x8 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 |
| Price? (> 20k) x9 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| Liked (y) | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 |
| **Prediction** | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |

1. Test Data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | phone11 | phone12 | phone13 | phone14 |
| Launch (within 6 months) x1 | 1 | 0 | 0 | 1 |
| Weight (<160g) x2 | 0 | 1 | 1 | 1 |
| Screen Size (< 5.9in) x3 | 0 | 1 | 1 | 1 |
| Dual sim x4 | 0 | 1 | 0 | 0 |
| Internal mem(>= 64gb, 4gb ram) x5 | 1 | 0 | 0 | 0 |
| NFC x6 | 0 | 0 | 1 | 0 |
| Radio x7 | 1 | 1 | 1 | 0 |
| Battery (>= 3500mAh) x8 | 1 | 1 | 1 | 0 |
| Price? (> 20k) x9 | 0 | 0 | 1 | 0 |
| Liked (y) | 0 | 1 | 0 | 0 |
| **Prediction** | 0 | 1 | 1 | 0 |

1. Accuracy = No. of correct predictions/ Total No. of predictions (¾ = 75% in test set)

## 2.1.6: Geometric Basics

1. Chapter on geometry basics, a brush-up.
2. x2 = mx1 + c
3. In 2D: General form ax1 + bx2 + c = 0
   1. Consider a = 2, b = 1, c = -2
   2. The intercepts are 1 and 2
   3. Consider the point (1,2), plugging it into the equation gives us the value 2
   4. If ax1 + bx2 + c > 0 then it is above the line
   5. If ax1 + bx2 + c < 0 then it is below the line
   6. If ax1 + bx2 + c = 0 then it is on the line
4. In 3D: General form ax1 + bx2 + cx3 + d = 0
   1. If ax1 + bx2 + cx3 + d > 0 then it is above the line
   2. If ax1 + bx2 + cx3 + d < 0 then it is below the line
   3. If ax1 + bx2 + cx3 + d = 0 then it is on the line

## 2.1.7: MP Neuron Geometric Interpretation

1. In 2D: ax1 + bx2 + d = 0
   1. x2 = -(a/b)x1 - (d/b)
   2. x2 = mx1 + c
   3. Where m = -a/b
   4. c = -d/b
2. = (ni=1xi >= b) in 2D can be rewritten as
   1. x1 + x2 - b >= 0 (decision boundary)
   2. Positive predictions(1) yield a value >= 0 and lie above the decision boundary
   3. Negative predictions(0) yield a value < 0 and lie below the decision boundary
3. This is a very restrictive model with respect to the freedom it has due to only one parameter
4. Some downsides to this model
   1. Boolean inputs and outputs
   2. The model is linear
   3. The model has a fixed slope
   4. The model has few possible intercepts(b’s)

## 2.1.8: Summary

1. Data: All boolean inputs ☹️
2. Task: Binary classification (boolean output) ☹️
3. Model: Linear decision boundary, all +ve points lie above the line ane -ve points are below (minimum flexibility) ☹️
4. cost/loss: mean squared error
5. Learning: brute force approach to learn best parameter b ☹️
6. Evaluation: Accuracy

# 2.2: Perceptron

Introduction to the Perceptron, a summary.

* Data: real inputs 😃
* Task: Classification(boolean output) ☹️
* Model: Weights for every input 😃, but still linear ☹️
* Cost/loss: imax(0,1 - yi \* **i)**☹️
* Learning: Our first learning algorithm 😃
* Evaluation: accuracy

## 2.2.1: Perceptron Data and Task

What kind of data and tasks can Perceptron process

1. Perceptron can also take real inputs

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | phone1 | phone2 | phone3 | phone4 | phone5 | phone6 | phone7 | phone8 | phone9 |
| Launch (within 6 months) x1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 |
| Weight (g) x2 | 151 | 180 | 160 | 205 | 162 | 182 | 138 | 185 | 170 |
| Screen Size (< 5.9in) x3 | 5.8 | 6.18 | 5.84 | 6.2 | 5.9 | 6.26 | 4.7 | 6.41 | 5.5 |
| Dual sim x4 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| Internal mem(>= 64gb, 4gb ram) x5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| NFC x6 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Radio x7 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |
| Battery (mAh) x8 | 3060 | 3500 | 3060 | 5000 | 3000 | 4000 | 1960 | 3700 | 3260 |
| Price? (k) x9 | 15k | 32k | 25k | 18k | 14k | 12k | 35k | 42k | 44k |
| Liked (y) | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 |

1. Apply feature scaling to standardize real input values x’ = (x-min)/(max-min)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Weight (g) x2 | 0.19 | 0.63 | 0.33 | 1.00 | 0.36 | 0.66 | 0.00 | 0.70 | 0.48 |
| Screen Size (< 5.9in) x3 | 0.64 | 0.87 | 0.67 | 0.88 | 0.70 | 0.91 | 0.00 | 1.00 | 0.47 |
| Battery (mAh) x8 | 0.36 | 0.51 | 0.36 | 1.00 | 0.34 | 0.67 | 0.00 | 0.57 | 0.43 |
| Price? (k) x9 | 0.09 | 0.63 | 0.41 | 0.19 | 0.06 | 0.00 | 0.72 | 0.94 | 1.00 |

## 2.2.2: Perceptron Model



1. = 1 if ni=1 wixi >= b
2. = 0 otherwise
3. Comparing with MP Neuron

|  |  |
| --- | --- |
| MP Neuron | Perceptron |
| = 1 if ni=1xi >= b  = 0 otherwise | = 1 if ni=1wixi >= b  = 0 otherwise |
| Boolean inputs ☹️ | Real inputs 😃 |
| Linear ☹️ | Linear ☹️ |
| Inputs are not weighted ☹️ | Weights for each input 😃 |
| Adjustable threshold 😃 | Adjustable threshold 😃 |

What do weights allow us to do?

1. Each parameter has a different effect on the output, some more, some less, some directly proportional and some inversely proportional.
2. Weights(/w) allow us to do this effectively.
3. x = [0, 0.19, 0.64, 1, 1, 0] features
4. w = [0.3, 0.4, -0.3, 0.1, 0.5] weights
5. x.w = ni=1 wixi
6. = 1 (if x.w >= b)
7. = 0 otherwise

## 2.2.3: Perceptron Geometric Interpretation

What is the geometric interpretation of the perceptron model

1. b can now take real values, and slope can change by varying w
2. x1 + x2 - b = 0
3. w1x1 + w2x2 - b = 0
4. x2  = -(w1/w2)x1 + (b/w2)
5. This results in more freedom than MP Neuron.
6. However, it only works with linearly separable data

## 2.2.4: Perceptron Loss Function

What loss function should you use for this model?

1. Consider the following training data

|  |  |  |
| --- | --- | --- |
| Weight | Screen Size | Liked(y) |
| 0.19 | 0.64 | 1 |
| 0.63 | 0.81 | 1 |
| 0.33 | 0.67 | 0 |
| 1 | 0.88 | 0 |

1. Loss/cost
   1. = 0 if y = ŷ,
   2. = 1 otherwise
2. More often, it is represented using an indicator variable
   1. L = 1(y ≠ ŷ)
   2. Or L = 0(y = ŷ)
3. **Q**: what is the purpose of the loss function
   1. **A**: It is to tell the model that some correction needs to be done
4. Comparing to Square Error loss function

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Weight | Screen Size | Liked(y) | ŷ | Perceptron loss  L = 1(y ≠ ŷ) | Sq. Error  (y - ŷ)2 |
| 0.19 | 0.64 | 1 | 0 | 0 | 0 |
| 0.63 | 0.81 | 1 | 0 | 1 | 1 |
| 0.33 | 0.67 | 0 | 1 | 1 | 1 |
| 1 | 0.88 | 0 | 0 | 0 | 0 |

1. The Perceptron loss is almost identical to the square error loss function. For all intents and purposes in this course, it can be considered equivalent to the square error loss function.

## 2.2.5: Perceptron Learning - General Recipe

What is the typical recipe for learning parameters of a model

1. Consider the following data

|  |  |  |
| --- | --- | --- |
| Weight  x1 | Screen Size  x2 | Liked(y) |
| 0.19 | 0.64 | 1 |
| 0.63 | 0.81 | 1 |
| 0.33 | 0.67 | 0 |
| 1 | 0.88 | 0 |

1. Randomly initialize parameters w1(𝜃1), w2(𝜃2) and b(𝜃0)
2. Iterate over data:
   1. L = compute\_loss(xi)
   2. update(w1,w2,b,L)
   3. Repeat till satisfied, till zero loss or some defined value ε is reached.

## 2.2.6: Perceptron Learning Algorithm

What does the perceptron learning algorithm look like?

1. Perceptron model: ŷ = ni=1 wixi >= b
   1. Can be rewritten as w1x1 + w2x2 - b >= 0
   2. Let w0 = b and x0 = 1
   3. Further rewritten as w1x1 + w2x2 - w0x0 >= 0
   4. ŷ = ni=0 wixi >= 0
   5. Can be written as wTx >= 0
   6. Where wTx = w.x
2. Perceptron Learning Algorithm
   1. P ⇒ Inputs with label 1
   2. N ⇒ Inputs with label 0
   3. Initialize w(w0...wn) randomly
   4. While !convergence do:
      1. Pick random x ∈ P ∪ N
      2. If x ∈ P and ni=0 wixi < 0 then, w = w + x; **end**
      3. If x ∈ N and ni=0 wixi >= 0 then, w = w - x; **end**
   5. **end**
   6. The algorithm converges when all the inputs are classified correctly

## 2.2.7: Perceptron Learning - Why it works?

What is the intuition behind the Perceptron learning algorithm

1. W = [w1, w2, … wn]
2. X = [x1, x2, … xn]
3. Cos 𝜃 = w.x/||w||x||, here the numerator can be replaced with wixi
   1. The denominator is always positive
   2. Therefore Cos 𝜃 ∝ wixi
   3. As 𝜃 ranges from 0 to 180o, cos 𝜃 ranges from 1 to -1
   4. If cos𝜃 > 0, it is an acute angle
   5. If cos𝜃 < 0, it is an obtuse angle
4. For x ∈ P, if w.x < 0, then it means that the angle(𝛂) between this x and the current w is greater than 90o, but we want 𝛂 < 90o)
   1. What happens to the new angle 𝛂new when wnew = w + x
   2. Cos𝛂new ∝ wnewTx
   3. ∝ (w+x)Tx
   4. ∝wTx + xTx(always +ve)
   5. ∝cos𝛂 + xTx(some +ve value)
   6. This means that the cosine is going to increase, which leads to decrease of 𝛂
5. For x ∈ N, if w.x > 0, then it means that the angle(𝛂) between this x and the current w is less than 90o, but we want 𝛂 > 90o)
   1. What happens to the new angle 𝛂new when wnew = w - x
   2. Cos𝛂new ∝ wnewTx
   3. ∝ (w-x)Tx
   4. ∝wTx - xTx(always +ve)
   5. ∝cos𝛂 - xTx(some +ve value)
   6. This means that the cosine is going to decrease, which leads to increase of 𝛂

## 2.2.8: Perceptron Learning - Will it always work

Will this algorithm always work?

1. It will only work if the data is linearly separable
2. If it is not linearly separable, the algorithm will never converge (ie, predict all training examples correctly)
3. Linearly Separable: Two sets P and N of points in an n-dimensional space are called absolutely linearly separable if
   1. n+1 real numbers wo,w1,...wn exist such that
   2. Every point(xo,x1,...xn) ∈ P satisfies ni=1 wixi >= w0
   3. Every point(xo,x1,...xn) ∈ N satisfies ni=1 wixi < w0
4. If the sets P and N are finite and linearly separable, the Perceptron learning algorithm will converge in a finite number of steps

## 2.2.9: Perceptron Evaluation

How do you check the performance of the perceptron model

1. Training data

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | phone1 | phone2 | phone3 | phone4 | phone5 | phone6 | phone7 | phone8 | phone9 |
| Launch (within 6 months) x1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 |
| Weight (g) x2 | 0.19 | 0.63 | 0.33 | 1.00 | 0.36 | 0.66 | 0.00 | 0.70 | 0.48 |
| Screen Size (< 5.9in) x3 | 0.64 | 0.87 | 0.67 | 0.88 | 0.70 | 0.91 | 0.00 | 1.00 | 0.47 |
| Dual sim x4 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| Internal mem(>= 64gb, 4gb ram) x5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| NFC x6 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Radio x7 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |
| Battery (mAh) x8 | 0.36 | 0.51 | 0.36 | 1.00 | 0.34 | 0.67 | 0.00 | 0.57 | 0.43 |
| Price? (k) x9 | 0.09 | 0.63 | 0.41 | 0.19 | 0.06 | 0.00 | 0.72 | 0.94 | 1.00 |
| Liked (y) | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 |

1. Test data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | phone 10 | phone 11 | phone 12 | phone 13 |
| Launch (within 6 months) x1 | 1 | 0 | 0 | 1 |
| Weight (g) x2 | 0.23 | 0.34 | 0.44 | 0.54 |
| Screen Size (< 5.9in) x3 | 0.74 | 0.93 | 0.34 | 0.42 |
| Dual sim x4 | 0 | 1 | 0 | 0 |
| Internal mem(>= 64gb, 4gb ram) x5 | 1 | 0 | 0 | 0 |
| NFC x6 | 0 | 0 | 1 | 0 |
| Radio x7 | 1 | 1 | 1 | 0 |
| Battery (mAh) x8 | 1 | 1 | 1 | 0 |
| Price? (k) x9 | 0 | 0 | 1 | 0 |
| Liked (y) | 0 | 1 | 0 | 0 |
| Prediction (ŷ) | 0 | 1 | 1 | 0 |

1. Accuracy = ¾ = 75%

## 2.2.10: Perceptron Summary

When will you use perceptron

1. Data: real inputs
2. Task: Classification (Boolean output)
3. Model: ni=0 wixi >= 0
4. Cost/loss: Σi 1(yi ≠ ŷi)
5. Learning Algorithm: Randomly assign and adjust w and b iteratively till convergence
6. Evaluation: Accuracy
7. How does this tie into final project
8. Perceptron can be used for image detection, to detect if

## 2.1.11: Perceptron: Toy Example

Perceptron Learning Algorithm in action

1. Dataset

|  |  |  |
| --- | --- | --- |
| x1 | x2 | y |
| -1 | -1 | 0 |
| -5 | -2.5 | 0 |
| -7.5 | 7.5 | 0 |
| 10 | 7.5 | 1 |
| -2.5 | 12.5 | 0 |
| 5 | 10 | 1 |
| 5 | 5 | 1 |

1. P ⇒ Green points, N ⇒ Red points
2. Decision boundary line is given by w1x1 + w2x2 - b >= 0
   1. Or x2 = -(w1/w2)x1 + (b/w2)
   2. Can be rewritten as x2 = mx + c
   3. Where m = -(w1/w2) and c = (b/w2)
3. Initialize w randomly
   1. w1 = 1.00, w2 = 1.00, b = 5.00 ⇒ m = -1.00, c = 5.00
4. The line is x2 = -x1 + 5
5. While !convergence do:
   1. Pick random x ∈ P ∪ N
   2. If x ∈ P and ni=0 wixi < 0 then, w = w + x; **end**
   3. If x ∈ N and ni=0 wixi >= 0 then, w = w - x; **end**
   4. Consider x = [x0;x1;x2] and w = [w0;w1;w2], where x0 bias term is always 1
   5. On the 5th training example, condition c isn’t satisfied, so we recalculate w = w - x
   6. w1 = 3.5, w2 = -11.5, b = 6.00 ⇒ m = 0.3, c = -0.52
   7. The line changes, causing a new error on the 6th training example, so we calculate w = w+x
   8. w1 = 8.5, w2 = -1.5, b = 5.00 ⇒ m = 5.67, c = -3.33
   9. The resulting line predicts all examples perfectly, thus convergence is reached