Homework 1 – Report

Mustafa Alper Sayan

S051674

Steps:

1. Split data into positive and negative
2. Find prior positive and prior negative samples
3. Find mean and variance for positive and negative samples
4. Calculate likelihood by assuming distribution is Gaussian (normal distribution), use mean and variance.
5. Find posterior of positive and negative samples
6. Determine threshold of decision based on determined parameters.
7. Determine estimator using posteriors (negative posterior and positive posterior)
8. Determine accuracy
9. Calculate total loss

Steps in Detail:

1. Split data into positive and negative

	age	result
0	26	Positive
1	26	Positive
2	29	Positive
3	28	Positive
4	24	Positive

	age	result
30	36	Negative
31	49	Negative
32	35	Negative
33	46	Negative
34	33	Negative

2. Find prior positive and prior negative samples
   Formula for priors:

$$\frac{\#\ of\ positive\ samples}{\#\ of\ samples} = P(C = Positive)$$

$$\frac{\#\ of\ negative\ samples}{\#\ of\ samples} = P(C = Negative)$$

Results from the homework:

```
prior positive train: 0.3333333333333333
prior negative test: 0.6666666666666666
prior positive test: 0.3333333333333333
prior negative test: 0.6666666666666666
```

3. Find mean and variance for positive and negative samples
   Formulas:

$$mean = \sum \frac{x_i}{n}$$

$$\sigma^2 = variance = \frac{\sum(x - \mu)^2}{n}$$

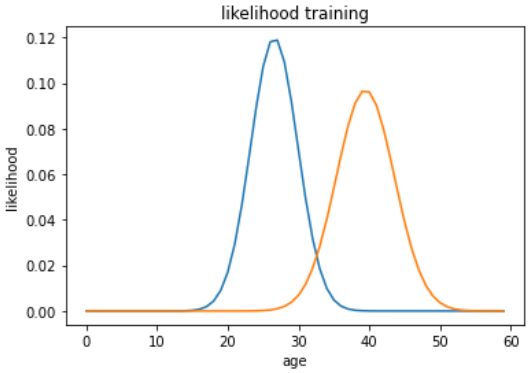
   Results from homework:

```
train mean positive: 26.566666666666666
train variance positive: 11.081609195402295
train mean negative: 39.45
train variance negative: 16.929661016949144
-----------------------------------------
test mean positive: 28.066666666666666
test variance positive: 10.616091954022984
test mean negative: 40.06666666666667
test variance negative: 18.842937853107344
```

4. Calculate likelihood by assuming distribution is Gaussian (normal distribution), use mean and variance. The output of the Gaussian distribution will be likelihoods. I assumed the ages could be between 0 – 60.

$$y = \frac{1}{\sigma\sqrt{2\pi}} * e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Results from the homework:



5. Find posterior of positive and negative samples

$$P(C = +|x = ?) = \frac{(Likelihood * Prior)}{Marginal\ Probability}$$

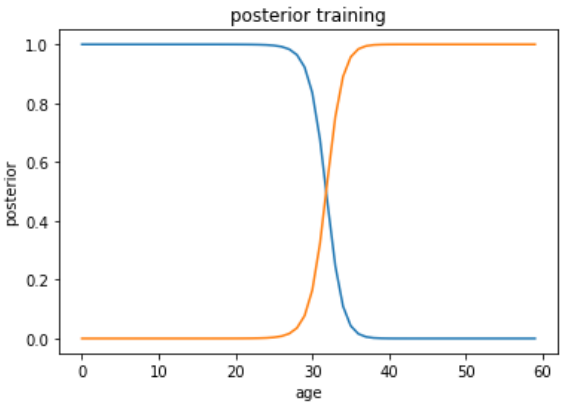
$$Marginal\ probility = P\ (x = ?\,|\,c = ?) * P(c = ?) + P(x = ?\,|\,c = ?) * P(c = ?)$$

Results from homework:
First 10 instances of posterior positive

```
[0.9999988051886746,
 0.9999988652203073,
 0.9999988881109589,
 0.9999988760447638,
 0.999998827874133,
 0.9999987389359253,
 0.9999986002911089,
 0.9999983972124998,
 0.999998106558812,
 0.9999976923745205]
```

Last 10 instances of posterior positive

[2.8735917889913763e-10,
6.366908139036632e-11,
1.3673963238170042e-11,
2.846574444897552e-12,
5.743981050681067e-13,
1.1234810850303779e-13,
2.1300060927018268e-14,
3.914337248588934e-15,
6.972651344328983e-16,
1.2039264967634808e-16]



6. Determine threshold of decision based on determined parameters.
   Formula:

$$R(\propto_1 | x) = loss_{c_1 \propto_1} * P(c_1 | x) + loss_{c_2 \propto_1} * P(c_2 | x)$$

$$R(\propto_2 | x) = loss_{c_1 \propto_2} * P(c_1 | x) + loss_{c_2 \propto_2} * P(c_2 | x)$$

   Choose $\propto_1$ (c_1) if

$$R(\propto_1 | x) < R(\propto_2 | x)$$

   Results from the homework:

   Loss of a False Positive = 1
   Loss of a False Negative = 1

$$R(\propto_1 | x) = 0 * P(c_1 | x) + 1 * P(c_2 | x) = P(c_2 | x)$$

$$R(\propto_2 | x) = 1 * P(c_1 | x) + 0 * P(c_2 | x) = P(c_1 | x)$$

$$P(c_2 | x) = 1 - P(c_1 | x)$$

$$1 - P(c_1 | x) < P(c_1 | x)$$

$$1 < 2P(c_1 | x)$$

$$0.5 < P(c_1 | x)$$

   Choose action 1 (positive class) if posterior of class 1 is bigger than 0.5

```
thresholds of decision for choosing a)
 action positive until posterior probility: 0.5
 action negative chosen after: 0.5

thresholds of decision for choosing b)
 action positive until posterior probility: 0.6666666666666666
 action negative chosen after: 0.33333333333333337

thresholds of decision for choosing c)
 action positive until posterior probility: 0.3333333333333333
 action negative chosen after: 0.6666666666666667
```

7. Determine estimator using posteriors (negative posterior and positive posterior)

   Assign Positive label until age ‘x’ if posterior possibility is bigger than the threshold of decision

   For homework:

```
age when first assigned negative for a) 32
age when first assigned negative for b) 32
age when first assigned negative for c) 33
```

8. Determine accuracy

```
training set accuracy of the predictor a): %94.44444444444444
False positive count train : 3 false negative count train: 2
 missclassified ages: [33, 32, 28, 31, 30]
----------------------------------------------------------------------
training set accuracy of the predictor b): %94.44444444444444
False positive count train : 3 false negative count train: 2
 missclassified ages: [33, 32, 28, 31, 30]
----------------------------------------------------------------------
training set accuracy of the predictor c): %94.44444444444444
False positive count train : 4 false negative count train: 1
 missclassified ages: [33, 28, 32, 31, 30]
----------------------------------------------------------------------
test set accuracy of the predictor a): %93.33333333333333
False positive count test : 3 false negative count test: 3
 missclassified ages: [36, 33, 33, 32, 32, 32]
----------------------------------------------------------------------
test set accuracy of the predictor b): %94.44444444444444
False positive count test : 0 false negative count test: 5
 missclassified ages: [36, 33, 33, 32, 32]
----------------------------------------------------------------------
test set accuracy of the predictor c): %94.44444444444444
False positive count test : 4 false negative count test: 1
 missclassified ages: [36, 32, 33, 32, 32]
----------------------------------------------------------------------
```

9. calculate total loss

$$(false\_positive\_count * loss\_false\_positive) + (false\_negative\_count * loss\_false\_negative)$$

For the homework:

```
total loss on train a) 5
total loss on train b) 8
total loss on train c) 6


total loss on test a) 6
total loss on test b) 5
total loss on test c) 6

total loss on test + train a) 11
total loss on test + train b) 13
total loss on test + train c) 12
```