Machine learning Exercise 3

Part 1 - Probability theory part

Q1 - MLE for Poisson.

Given a random sample $\{x1, x2, \dots, xn\}$, derive the maximum likelihood estimator λ of the Poisson distribution.

$$P\left(x, \lambda\right) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

We will look at $L(\theta) = lnP(D|\theta)$:

$$L(\theta) =$$
$$= lnP(D|\theta) =$$
$$= ln\left(\Pi_{j=1}^{n} e^{-\lambda}\frac{\lambda^{X_j}}{x_j!}\right) =$$
$$= \Sigma_{j=1}^{n} ln\left(e^{-\lambda}\frac{\lambda^{X_j}}{x_j!}\right) =$$
$$= \Sigma_{j=1}^{n}\left(ln(e^{-\lambda}) + ln(\lambda^{x_j}) - ln(X_j!)\right) =$$
$$= \Sigma_{j=1}^{n}\left(-\lambda + x_j ln\lambda - ln(x_j!)\right) =$$
$$= -n\lambda + ln\lambda\Sigma_{j=1}^{n}x_j - \Sigma_{j=1}^{n}ln(x_j!)$$

$$\frac{dL(\theta)}{d\lambda} = -n + \frac{1}{\lambda}\Sigma_{j=1}^{n}x_j = 0 \quad \Rightarrow \quad \lambda = \frac{1}{n}\Sigma_{j=1}^{n}x_j$$

Q2 - A radar at the beach is used to detect ships.

Ships are located in one of four zones: A, B, C and D.
The probability of detection per zone is 0.75, 0.5, 0.3, 0.4 for A, B, C,D respectively.
The probability of being at a specific zone is 0.4, 0.2, 0.3, 0.1 for A, B, C,D, respectively.

a. What is the probability that a ship will be detected?

$$P(detected) = P(detected|A)P(A) + P(detected|B)P(B) + P(detected|C)P(C) + P(detected|D)P(D) =$$
$$= 0.75 \cdot 0.4 + 0.5 \cdot 0.2 + 0.3 \cdot 0.3 + 0.4 \cdot 0.1 = 0.53$$

b. Given that a ship is detected, what is the probability that it was in zone C?

$$P(C|detected) = \frac{P(detected|C)P(C)}{P(detected)} = \frac{0.3 \cdot 0.3}{0.53} = 0.17$$

c. Given that a ship is detected, what is the probability that it was in zone B?

$$P(B|detected) = \frac{P(detected|B)P(B)}{P(detected)} = \frac{0.5 \cdot 0.2}{0.53} = 0.189$$

Q3 - Find 3 random variables X, Y, C such that:

a. $X \perp Y \mid C$ – (X and Y are independent given C.)
b. X and Y are not independent.
c. X, Y take integer values such that $1 \leq X, Y \leq 10$ and C is binary.
d. The following conditions hold:

    i. $P(1 \leq X \leq 5) = 0.3$
    ii. $P(1 \leq Y \leq 5) = 0.3$
    iii. $P(C = 0) = 0.5$

Let's define `C` as a uniform binary variable:

$$C \sim U(0, 1)$$
$$P(C = 0) = 0.5$$
$$P(C = 1) = 0.5$$

Let's Define X as follows:

$X \in \{1, 2, 3...10\}$

$P\big(1 \le X \le 10 \mid C = 0\big) = 0$

$P\big(1 \le X \le 5 \mid C = 1\big) = 0.12$

$P\big(6 \le X \le 9 \mid C = 1\big) = 0.2$

$P\big(X = 10 \mid C = 1\big) = 0.6$

Let's Define Y as follows:

$Y \in \{1, 2, 3...10\}$

$P\big(1 \le Y \le 10 \mid C = 1\big) = 0$

$P\big(1 \le Y \le 5 \mid C = 0\big) = 0.12$

$P\big(6 \le Y \le 9 \mid C = 0\big) = 0.2$

$P\big(Y = 10 \mid C = 0\big) = 0.6$

Thus X and Y are dependent(not independent)!
Because, For example, $P(X = 10) < Pr(X = 10|Y = 10)$.

Keepiing with the conditions:

$$P\big(1 \le X \le 5\big) \;=\; P\big(1 \le X \le 5 \mid C = 0\big) \cdot P(C = 0) + P\big(1 \le X \le 5 \mid C = 1\big) \cdot P(C = 1) \;=\; 0 \cdot 0.5 + (0.12 \cdot 5) \cdot 0.5 \;=\; 0.3$$

$$P\big(1 \le Y \le 5\big) \;=\; P\big(1 \le Y \le 5 \mid C = 0\big) \cdot P(C = 0) + P\big(1 \le Y \le 5 \mid C = 1\big) \cdot P(C = 1) \;=\; (0.12 \cdot 5) \cdot 0.5 + 0 \cdot 0.5 \;=\; 0.3$$

$$P\big(C = 0\big) \;=\; 0.5$$

Done!

Q4 - The probability of having a decent meal in Karnaf is 0.65.

a. What is the probability of having 3 descent meals in a week (5 days)?

$$P(x=3)=\binom{5}{3}0.65^3\cdot 0.35^2=0.34$$

b. What is the probability of having at least 2 descent meals in a week?

$$P(2\le x)=$$
$$=1-\Big(P(x=0)+P(x=1)\Big)=$$
$$=1-\left(\binom{5}{0}0.65^0\cdot 0.35^5+\binom{5}{1}0.65^1\cdot 0.35^4\right)=$$
$$=1-\big(0.0053+0.049\big)=0.946$$

c. A class of 300 students recorded the number of descent meals they had during a specific week.

They averaged their results. What do you expect the value of that average to be?

let's note `X` as total number of decent meals.

$X \sim Bin(5, 0.65)$

Average of Number of descemt meals = $\frac{Student_1 + student_2 + ... + Student_{300}}{300} =$

And as we make our sample size larger it should be close to the mean for a single sample which is

$\Rightarrow E(X) = n \cdot p = 5 \cdot 0.65 = 3.25$

Q5 - Bivariate Normal Distribution

you are given a dataset of 1,000 (x1, x2) points drawn from a Bivariate Normal distribution with unknown parameters (data/bivariate_normal_data.csv).

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
from scipy.stats import multivariate_normal
```

a. Estimate the distribution parameters using the following (these are the MLE parameters):

$$\mu_i = \frac{1}{N}\sum_{k=1}^{N} x_i^{(k)}$$

$$\sigma_i = \frac{1}{N}\sum_{k=1}^{N} \left(x_i^{(k)} - \mu\right)^2$$

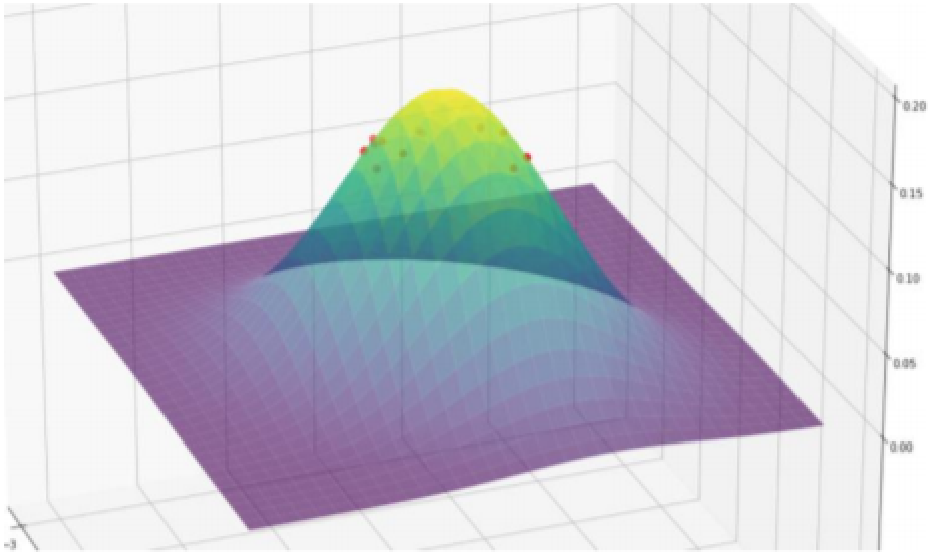
$$\rho = \frac{1}{N}\sum_{k=1}^{N} \left(x_1^{(k)} - \mu_1\right) \cdot \left(x_2^{(k)} - \mu_2\right)$$

In [2]:
```
df = pd.read_csv('data/bivariate_normal_data.csv',names=['x1','x2'],header=0)
df.head()

μ = df.mean() # or df.sum(axis=0)/len(df)
σ = np.square(df - μ).mean()  # or df.std()
ρ = ((df.x1 - μ[0]) * (df.x2 - μ[1])).mean()

print('μ = ', μ.values)
print('σ = ', σ.values)
print('ρ = ', ρ)
```

```
μ =  [ 0.51113139 -2.01599143]
σ =  [0.99302132 0.99689616]
ρ =  0.6619576045042657
```

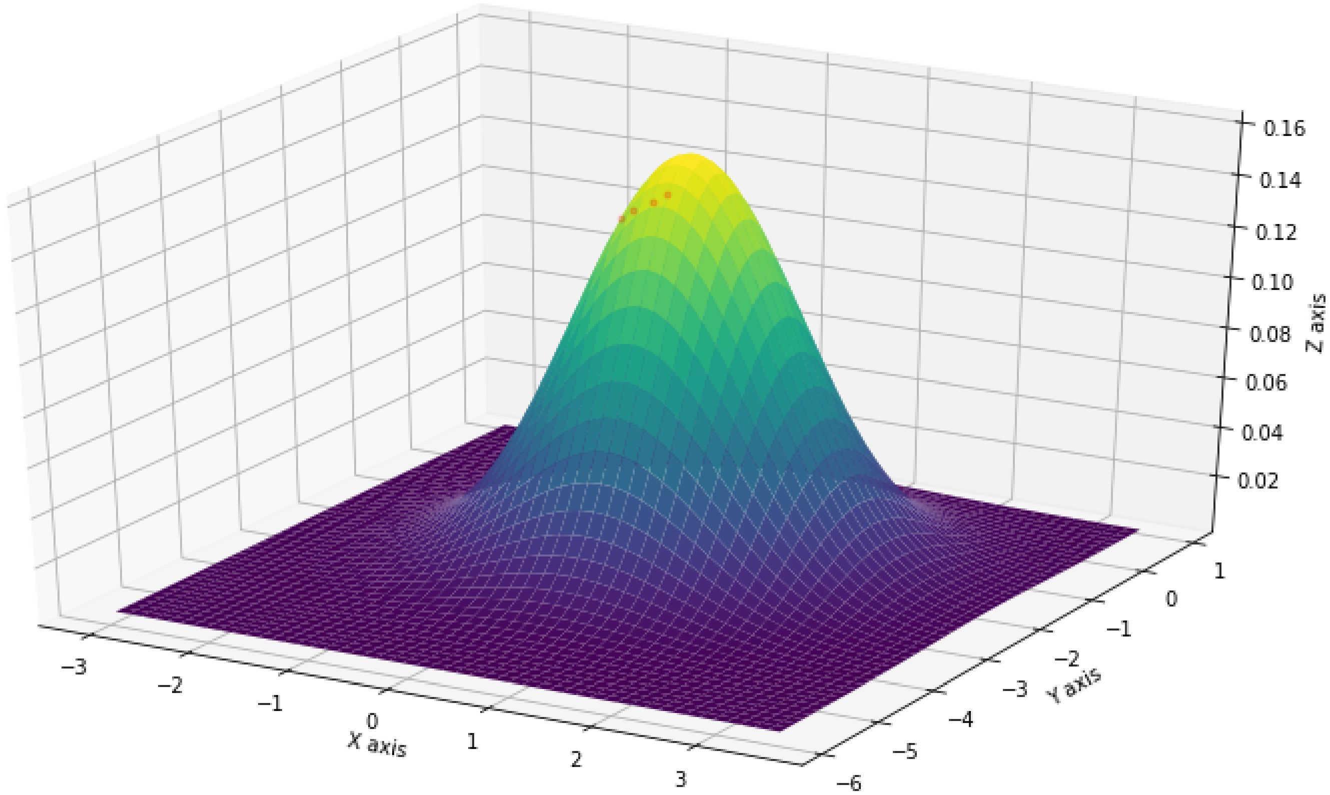
b. Using the parameters you found:

- plot the distribution in 3d
- mark on the plot the points from the dataset we provided you that correspond to some narrow equiprobable range.
- For example, given the range $P(0.14 \le X \le 0.15)$:



In [3]:

```
# Sort values
x = df.x1.sort_values()
y = df.x2.sort_values()


# Surface - Create grid and multivariate normal
X, Y = np.meshgrid(np.linspace(x.min(),x.max(),500),np.linspace(y.min(),y.max(),500))
pos = np.empty(X.shape + (2,))
pos[:, :, 0] = X
pos[:, :, 1] = Y
Z = multivariate_normal(μ.values, np.square(σ.values)).pdf(pos)

# points - Create z values by range
x_min,x_max = 0.14,0.15
xy = np.empty( x.shape + (2,) )
xy[:,0] = x
xy[:,1] = y
z = multivariate_normal(μ.values, np.square(σ.values)).pdf(xy)
x_range = (x_min <= x) & (x <= x_max)

#Make a 3D plot
fig = plt.figure(figsize=(14,8))
ax = fig.gca(projection='3d')
ax.plot_surface(X, Y, Z,cmap='viridis',linewidth=0)
ax.plot(x[x_range], y[x_range], z[x_range], marker='.', linestyle='None', label='Label', color='red', zorder=10,alpha=
0.2)
ax.set_xlabel('X axis')
ax.set_ylabel('Y axis')
ax.set_zlabel('Z axis')
# ax.view_init(elev=30,azim=200) # change view of the 3D plot
plt.show()
```


0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
Z axis
−3
−2
−1
0
1
2
3
X axis
1
0
−1
−2
−3
−4
−5
−6
Y axis