1) k-Means Clustering

a) The data:

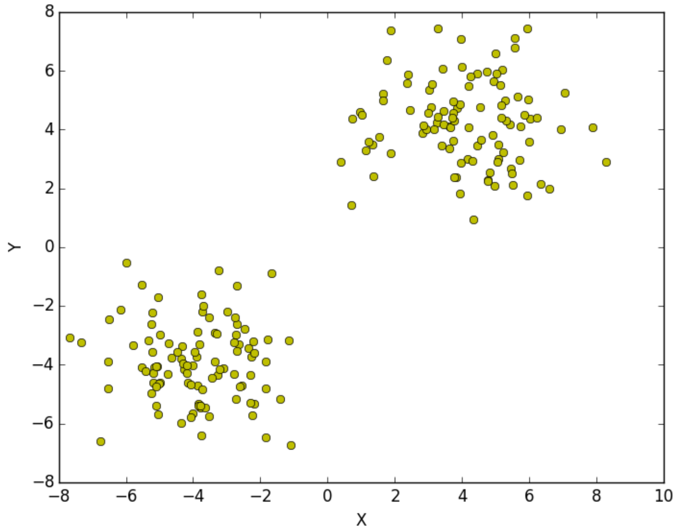

Figure 1: Data points

b) Implementation of K-means (see **k-means.py**):
c) The clusters:

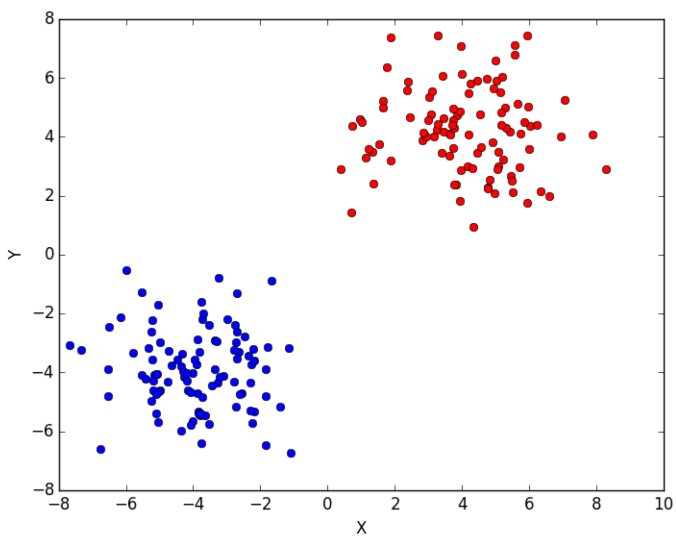

Figure 2: Cluters

d) To get a loss function equals to zero, the distance of each data point to its mean must be equal to zero. In particular, the data point is exactly the same as its mean. So, the number k is at least equal to number of **distinct** data points.

2) Maximum Likelihood Estimation

We have trained data:

$$K = \{2, 3, 0, 2, 1, 5\}$$

We need to maximize the likelihood of trained data:

$$\begin{aligned}
&\max \prod_i \frac{\lambda^{k_i} e^{-\lambda}}{k_i!} \\
\equiv\ &\max \prod_i \lambda^{k_i} e^{-\lambda} \\
\equiv\ &\max \lambda^{\sum_i k_i} e^{-\lambda|K|} \\
\equiv\ &\max \lambda^{13} e^{-6\lambda}
\end{aligned}$$

Take the derivative:

$$\begin{aligned}
(\lambda^{13} e^{-6\lambda})' &= (13\lambda^{12} e^{-6\lambda}) + (\lambda^{13}(-6)e^{-6\lambda}) \\
&= (\lambda^{12} e^{-6\lambda})(13 - 6\lambda)
\end{aligned}$$

Let the derivative be zero:

$$(\lambda^{12} e^{-6\lambda})(13 - 6\lambda) = 0 \iff \lambda = 0 \; OR \; \lambda = \frac{13}{6}$$

Because λ could not be zero, so we have $\lambda = \frac{13}{6}$

3) Composite functions

$$f(x, y) = log(sin(xy))$$

First order derivative with respect to x:

$$\begin{aligned}
\frac{\partial f}{\partial x} &= \frac{\partial(log(sin(xy)))}{\partial x} \\
&= \frac{1}{sin(xy)} \frac{\partial(sin(xy))}{\partial x} \\
&= \frac{1}{sin(xy)} cos(xy) \frac{\partial(xy)}{\partial x} \\
&= y \frac{cos(xy)}{sin(xy)} = y \cot(xy)
\end{aligned}$$

The same, we have first order derivative with respect to y:

$$\frac{\partial f}{\partial y} = x \cot(xy)$$

Second order partial derivative with respect to x and y:

$$\begin{aligned}
\frac{\partial^2 f}{\partial xy} = \frac{\partial^2 f}{\partial yx} &= \frac{\partial(y \cot(xy))}{\partial y} \\
&= cot(xy) \frac{\partial y}{\partial y} + y \frac{\partial(cot(xy))}{\partial y} \\
&= cot(xy) + y(-\frac{x}{sin^2(xy)}) = cot(xy) - \frac{xy}{sin^2(xy)}
\end{aligned}$$

Second order partial derivative with respect to x:

$$\begin{aligned}\frac{\partial^2 f}{\partial x^2} &= \frac{\partial(y\cot(xy))}{\partial x} \\ &= cot(xy)\frac{\partial y}{\partial x} + y\frac{\partial(cot(xy))}{\partial x} \\ &= y(-\frac{y}{sin^2(xy)}) = -\frac{y^2}{sin^2(xy)}\end{aligned}$$

The same, we have second order partial derivative with respect to y:

$$\frac{\partial^2 f}{\partial y^2} = -\frac{x^2}{sin^2(xy)}$$

4) Classification

See **logistic_regression.py**

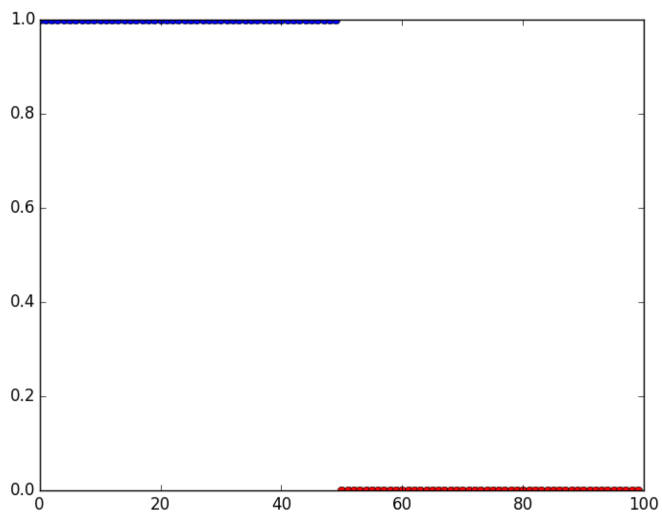

Figure 3: Classification Result