**1.Binary classification**

- output is 0 or 1, NG or OK,...

- input is image -> convert to arr include all pixel value

Example:image(64x64) -> input is vector 3x64x64 pixel value

**2.Logistic Regression**

- input is X, want output y^ = P(y=1|x) (0<y^<1)

- parameter is w and b

- linear regression function: y=wx+b can not use in here because output of logistic regression is (0,1)

so y^ = f(W^Tx+b) = f(z) = 1/(1+e^-z) (Sigmod function have output in (0,1)

**3.Loss function & Cost function**

- Loss function:

Loss function is function compute different between y^ and y of 1 sample data

Given (x1,y1);(x2,y2);....(xm,ym)

we want to y^(i) ~ y(i) with y^(i)=f(z)=f(W^Tx+b)

So Loss function: L = 1/2(y^-y)^2. This loss function suitable for linear regression

In logistic regression, we use different loss function: L = -(ylog(y^) +(1-y)log(1-y^))

This function suitable when y~1 then y^ large, when y~0 then y^ small

- Cost function:

Cost function is loss function when compute with m sample data:

C(w,b) = 1/m ∑[-(y(i)log(y^(i)) +(1-y(i))log(1-y^(i)))]

**4.Gradient descent**

Gradient descent is a method find minimize value of function.

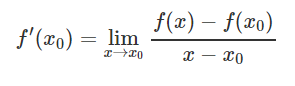
Example y = f(x), so what is the x make y minimize?

we have update: x(t+1)=x(t)−nf′(x(t)) to find x make y minimize

So gradient descent include 2 part importance: +) Begin point

+) Learning rate (n)

**5. Derivative**



**6. Logistic Regression Derivatives**

Target: find (w,b) make cost function of logistic regression is minimize

