8. Classification with logistic regression.

I. Motivations.

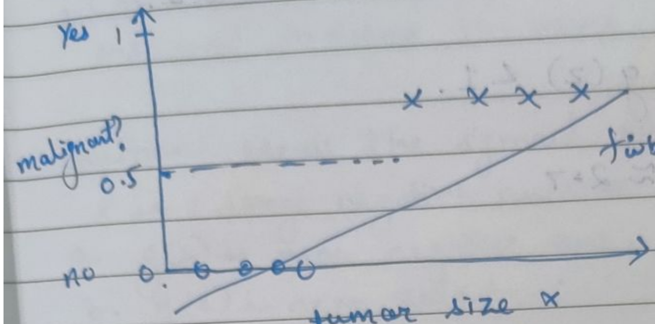
Classification Problems.

Question?	Answer?
is email spam?	→ no or yes
is the txn fraudulent?	→ no or yes.
is the tumor malignant	→ no or yes.

y can only be one of two values.

This is called 'binary' classification. class = category. We can have flase → 0, true = 1. This is useful for classification.

flase → 0 → negative class. the absence.
true → 1 - positive class. the presence.



Can we apply linear regression?.

$f_{w,b}(x) = wx + b$.

We can try picking a threshold.

if $f_{w,b}(x) < 0.5 \rightarrow \hat{y} = 0$
if $f_{w,b}(x) \geq 0.5 \rightarrow \hat{y} = 1$

} fo this ~~set~~ set, our example works for our dataset.

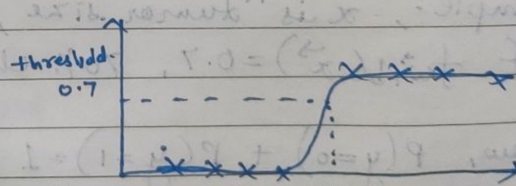
adding a new tumor size at the right, our learning algo becomes much worse. It starts misclassifying.

This threshold is called the decision boundary.

Question: Which of the following is a classification task?.
a) Decide if an animal is a cat or not a cat.
b) Estimate the weight of a cat based on it's height.

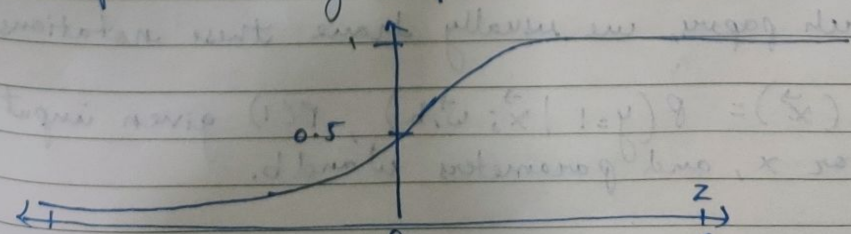
Answer - a.

II. Logistic Regression.
Continuing with our example.



We want output between 0 and 1.

Sigmoid function - Logistic function



sigmoid function.

Sigmoid function outputs values between 0 and 1.

$g(z) = \frac{1}{1+e^{-z}}$. $0 < g(z) < 1$.

$e \approx 2.7$

if z is ≈ 100, $g(z) \approx 1$
if z is ≈ -100, $g(z) \approx 0$.
if $z = 0$, $g(z) = 0.5$

We remember our, $f_{\vec{w},b}(\vec{x}) = \vec{w} \cdot \vec{x} + b$.

Let $\vec{z} = \vec{w} \cdot \vec{x} + b$, now $g(z) = \frac{1}{1+e^{-z}}$, now our

$0 \leq g(z) \leq 1$.

$$f_{\vec{w},b}(\vec{x}) = g(\vec{w} \cdot \vec{x} + b) = \frac{1}{1+e^{-(\vec{w} \cdot \vec{x} + b)}}$$

This is basically logistic regression.

We basically calculate the probability that the class is 1.

Example:, x is "tumor size", y is 0 or 1.
if $f_{\vec{w},b}(\vec{x}) = 0.7$, 70% chance that y is 1.

Now, $P(y=0) + P(y=1) = 1$ (Probabilities).
so $P(y=0) = 1 - P(y=1) = 1 - 0.7 = 0.3$
30% chance that y is 0.

In research papers, we usually have these notations.

$f_{\vec{w},b}(\vec{x}) = P(y=1 \mid \vec{x}; \vec{w}, b)$, $P(1)$ given input vector x, and parameters $\vec{w}$ and b.

; (semicolon) here denotes that $\vec{w}, b$ are parameters, where are used to define the value of the probability.

Question: Recall the sigmoid function, $g(z) = \frac{1}{1+e^{-z}}$, if z is a large negative number then:

a. g(z) is near negative one.

b. g(z) is near zero.

Answer - b.

III. Decision Boundary.

$f_{\vec{w},b}(\vec{x})$ *, we first calculate $z = \vec{w} \cdot \vec{x} + b$.

then $g(z) = \frac{1}{1+e^{-z}}$.

$$f_{\vec{w},b}(\vec{x}) = g(\vec{w}\cdot\vec{x}+b) = \frac{1}{1+e^{-z(\vec{w}\cdot\vec{x}+b)}} = \frac{1}{1+e^{-(\vec{w}\cdot\vec{x}+b)}}$$

$$= P(y=1 \mid \vec{x}; \vec{w}, b)$$

0 or 1?.

Is $f_{\vec{w},b}(\vec{x}) \geq 0.5$? $\rightarrow \hat{y} = 1$

$f_{\vec{w},b}(\vec{x}) < 0.5 \rightarrow \hat{y} = 0$.

Now, when is $f_{\vec{w},b}(\vec{x}) \geq 0.5$?

Now, when is $g(z) \geq 0.5$,

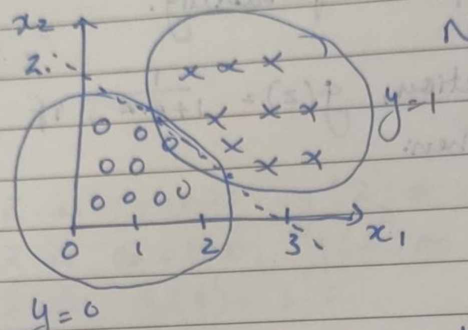
whenever $z \geq 0$,

$\vec{w}\cdot\vec{x} + b \geq 0$, we predict $\hat{y} = 1$

$\vec{w}\cdot\vec{x} + b < 0$, we predict $\hat{y} = 0$

} based on sigmoid function.

b. Decision boundary.



Now, $f_{\vec{w},b}(\vec{x}) = g(z)$

$= g(w_1 x_1 + w_2 x_2 + b)$

$z = \vec{w} \cdot \vec{x} + b = 0.$

Now, this decision boundary, $z = \vec{w} \cdot \vec{x} + b - 0$

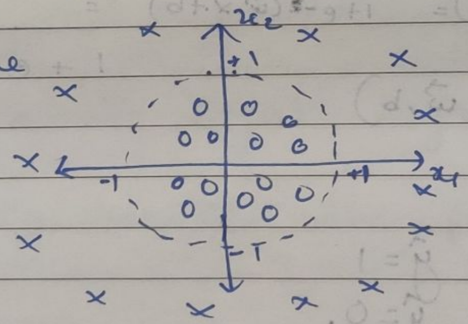
$z = x_1 + x_2 - 3$ } for $w_1 = 1,\ w_2 = 1,\ b = -3.$

When $z=0$; $x_1 + x_2 - 3 = 0 \rightarrow x_1 + x_2 = 3$

This line can be seen as the dotted line, which is our decision boundary for logistic regression.

For different parameters, we have different boundaries.

If we have



Now, $f_{\vec{w},b}(\vec{x}) = g(z) = g(w_1 x_1^2 + w_2 x_2^2 + b)$

Let $w_1 = 1,\ w_2 = 1,\ b = -1$

So, $z = x_1^2 + x_2^2 - 1$,

decision boundary is $z = 0 \rightarrow x_1^2 + x_2^2 - 1 = 0$

$x_1^2 + x_2^2 = 1$ } which is denoted in the dotted circle.

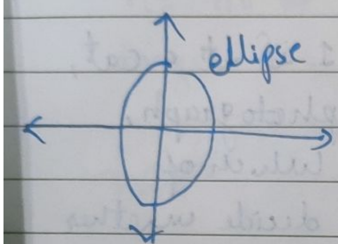
Now, when $x_1^2 + x_2^2 \geq 1$ $\hat{y} = 1$

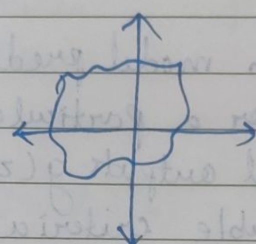
$x_1^2 + x_2^2 < 1$ $\hat{y} = 0.$

C. Non-Linear decision boundaries.

We can have complex boundaries as well.

say: $f_{\vec{w},b}(\vec{x}) = g(z) = g(w_1 x_1 + w_2 x_2 + w_3 x_1^2 + w_4 x_1 x_2 + w_5 x_2^2 \ldots + b)$.





complex boundaries.

If we don't use exponents, our decision boundary will always be a straight line.

Question: Let's say we are creating a tumor detection algorithm. Our algorithm will be used to flag potential tumors for future inspection by a specialist. What values should you use for a threshold?

a) High, say a threshold 0.9?

b) Low, say a threshold of 0.2?

Answer - b. Not missing potential tumor.

IV. Quiz.

1. Which is a classification task?

a) Based on patient's blood pressure, determine how much blood pressure medication.

b) Based on age and blood, determine how much blood pressure medication.

c) Based on size of tumor, determine if each tumor is malignant or not.

Answer - C.

2. Recall the sigmoid function $g(z) = \frac{1}{1+e^{-z}}$
if z is large positive number, then:
a) g(z) will be 0. b. will be -1 c) will be 1.
d) will be 0.5

Answer - C.

3. A cat photo classification model predicts 1 if it a cat, and 0 if it's not a cat. For a particular photograph, the logistic regression model outputs g(z). Which of these would be reasonable criteria to decide whether to predict if it's a cat?
a) Predict cat if $g(z) = 0.5$
b) cat if $g(z) \geq 0.5$ c) cat if $g(z) < 0.5$
d) cat if $g(z) < 0.7$.
Answer → b.

4. No matter what features you use, the decision boundary learned by logistic regression will be a linear decision boundary.
→ False.