Supervised ML

Institute of Technology of Cambodia

December 5, 2023

Classification

Classification

- Objective: predict the class of an item
- Methods for regression can be reused with some adaptations
  1. Binary classification is usually simple
  2. Multiclass classification might require more changes
- Evaluation is different

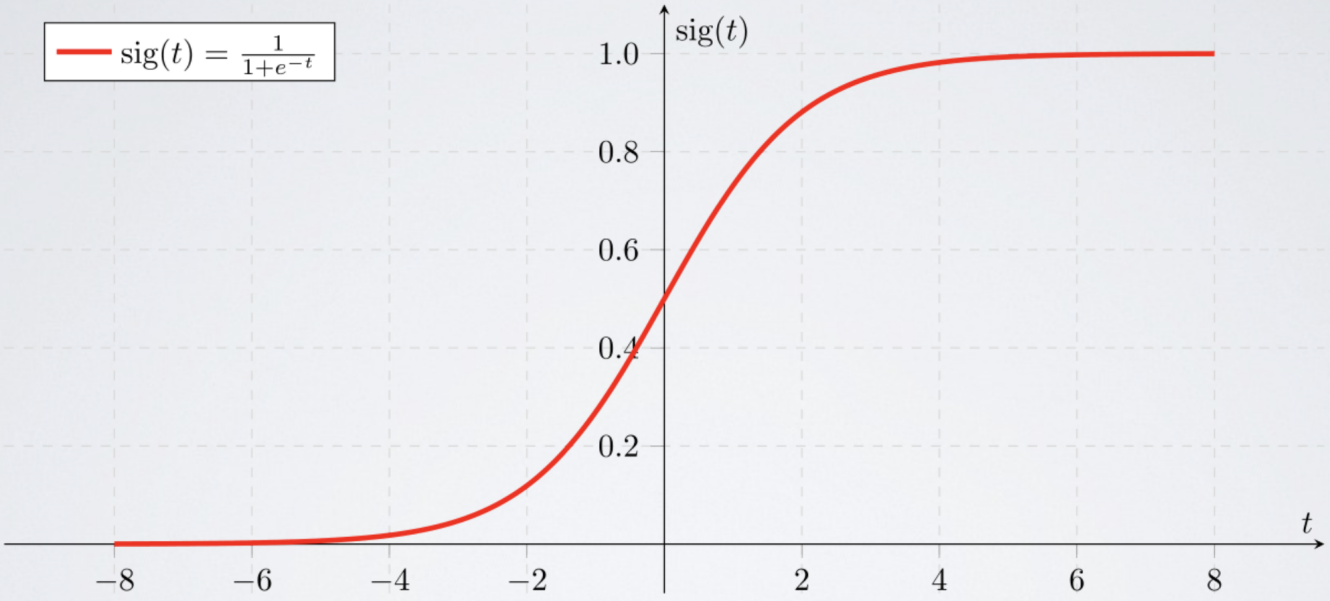
Logistic Regression

- Logistic Regression (also called Logit Regression) is commonly used to estimate the probability that an instance belongs to a particular class.
- Logistic Regression measures the relationship between the categorical dependent variable and one or more independent variables by estimating probabilities using a logistic function.
- `https://en.wikipedia.org/wiki/Logistic_regression`

Logistic Regression

How does it work?
Just like a Linear Regression model, a Logistic Regression model computes a weighted sum of the input features (plus a bias term), but instead of outputting the result directly like the Linear Regression model does, it outputs the logistic of this result.

Sigmoid function



$$\lim_{t \to -\infty} sig(t) = 0 \qquad \lim_{t \to +\infty} sig(t) = 1 \qquad sig(0) = 0.5$$

Logistic Regression

$$\text{Logistic (Sigmoid) function:} Sig(x) = \frac{1}{1 + e^{-x}}$$

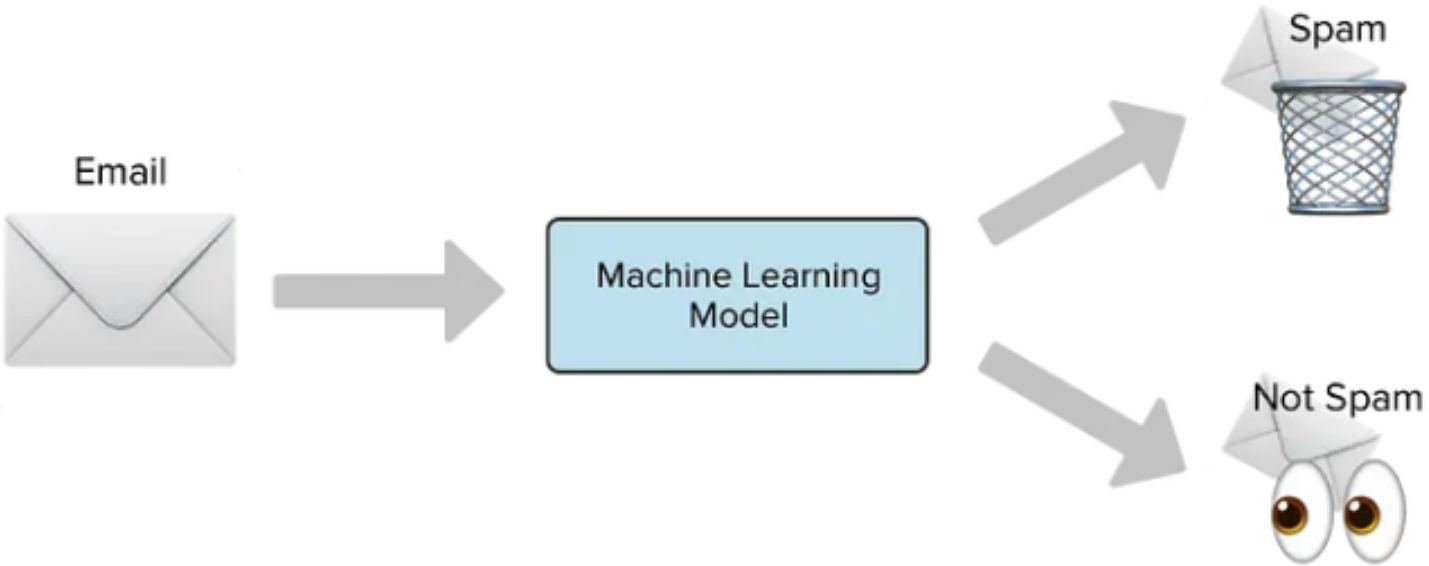
$$\text{Linear regression: } y' = \beta_0 + \beta_1 x_i + \beta_2 x_2 + \ldots + \beta_n x_n$$

$$\text{Logistic regression } P(y = 1) = Sig(\beta_0 + \beta_1 x_i) + \beta_{x2} x_2 + \ldots + \beta_n x_n)$$

$$\text{OR Logistic regression } P(y = 1) = \frac{1}{1 + e^{-\beta_0 + \beta_1 x_i + \beta_2 x_2 + \ldots + \beta_n x_n}}$$

Logistic Regression

Suppose we want to detect the spam email.



Logistic Regression

ham	Go until jurong point, crazy.. Available only in bugis n great world la e buffet... Cine there got amore wat...
ham	Ok lar... Joking wif u oni...
spam	Free entry in 2 a wkly comp to win FA Cup final tkts 21st May 2005. Text FA to 87121 to receive entry question(std txt rate)T&C's apply 08452810075over18's
ham	U dun say so early hor... U c already then say...
ham	Nah I don't think he goes to usf, he lives around here though
spam	FreeMsg Hey there darling it's been 3 week's now and no word back! I'd like some fun you up for it still? Tb ok! XxX std chgs to send, å£1.50 to rcv
ham	Even my brother is not like to speak with me. They treat me like aids patent.
ham	As per your request 'Melle Melle (Oru Minnaminunginte Nurungu Vettam)' has been set as your callertune for all Callers. Press *9 to copy your friends Callertune
spam	WINNER!! As a valued network customer you have been selected to receivea å£900 prize reward! To claim call 09061701461. Claim code KL341. Valid 12 hours only.
spam	Had your mobile 11 months or more? U R entitled to Update to the latest colour mobiles with camera for Free! Call The Mobile Update Co FREE on 08002986030

Logistic Regression

What is the probability that this email is spam?

- If the estimated probability is greater than 50%, then the model predicts that the instance belongs to that class (called the positive class, label "1")
- Else it predicts that it does not (i.e., it belongs to the negative class, labeled "0")

Multiclass Logistic Regression

In many cases, we have more than 2 classes

- e.g,: house, apartment, office, industrial, cat, dog, horse
- Categories are unordered, conversion to numeric would be catastrophic

Simple solution: one VS all

- Train a logistic classifier on one class VS all other classes
- Pick the class with the largest confidence
  - e.g.: house:20%, apartment:30%, office:70%, Industrial:80%
  (Answer: Industrial)

Alternative approach: softmax regression

SoftMax

Softmax is a generalization of Logistic/Sigmoid to Multiclass

- Takes several outputs with arbitrary values $\in (-\infty, +\infty)$.
- Convert into a set of (positive) probabilities summing to 1

$$\sigma(Z)_i = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}}$$

- Z: vector of real numbers
- Exponential convert Real into $(0,+\infty)$
- Division by the sum normalizes (sum of values $= 1$)

Cross Entropy

The usual loss function associated with softmax is the cross-entropy

- We have an estimated probability q(i) for each possible outcome i, we compare with the true distribution (one-hot encoding, p(i)=1 for the true label)

$$H(P,Q) = -\sum_i p(i) log q(i)$$

$$H(p,Q) = -log q(i) \text{for i=True label}$$

Classification evaluation

Binary Classification

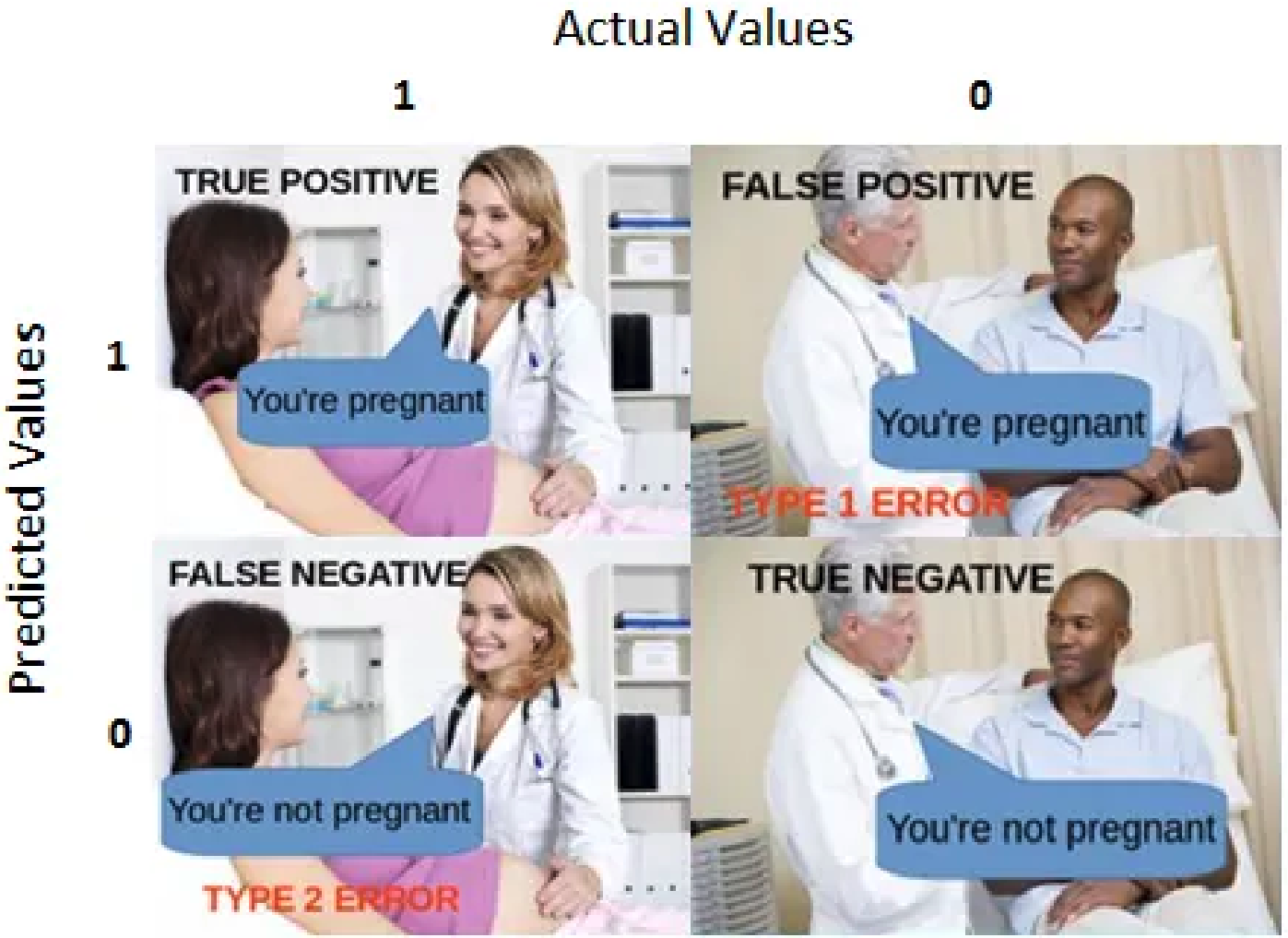
Many scenarios require binary classification

- Covid/not Covid
- Give a credit/do not give credit
- Spam/not-spam
- Positive sentiment/negative sentiment
- Face on a photo/no face
- Normal user/bot
- Etc

Evaluation

		Actual	
		Positive	Negative
Predicted	Positive	**True Positive**	**False Positive**
	Negative	**False Negative**	**True Negative**

Classification Evaluation



Classification Evaluation

$$\text{Precision} = \frac{TP}{TP + FP}$$

Among those predicted as True, fraction of really True

$$\text{Recall} = \frac{TP}{TP + FN}$$

Among those really true, what fraction did we identity correctly

None-symmetric

- Precision success != Precision failure

Accuracy

$$\text{Accuracy} = \frac{TP + TN}{P + N}$$

Fraction of correct prediction, among all predictions

- Simple to interpret, symmetric

Main drawback: class imbalance

- Test whole city, 1000 people, for Covid
  - 95% don't have covid, i.e, 50 people have covid, 950 don't have it
- Our test (ML algorithm) is pretty good: TP:45, FN:5, TN:900, FP:50
  - Accuracy=(45+900)/1000 = 0.945
- Dumb classifier: always answer: not covid
  - Accuracy: (0+950)/1000 = 0.95

F1 Score

$$\text{F1 score: } F_1 = 2\frac{precision * recall}{precision + recall}$$

Harmonic mean between precision and recall

- Harmonic mean more adapted for rates
- Give more importance to the lower value
- Not symmetric

Scores for the covid preditor:

- Precision=45/95=0.47
- Recall=45/50=0.9
- F1=0.65

Score for the naive predictor impossible to compute...

- You need at least some TP!
- Assuming 1 "free" TP (Precision=1, Recall=1/50)
- F1=0.04