Day 5: Mini-Project

Summer STEM: Machine Learning

Department of Electrical and Computer Engineering
NYU Tandon School of Engineering
Brooklyn, New York

June 26, 2020

Outline

Classification Vs. Regression

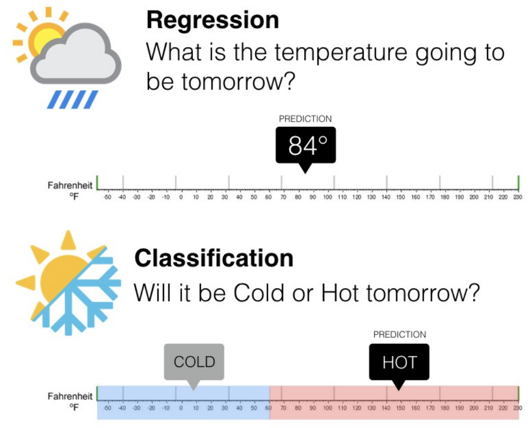
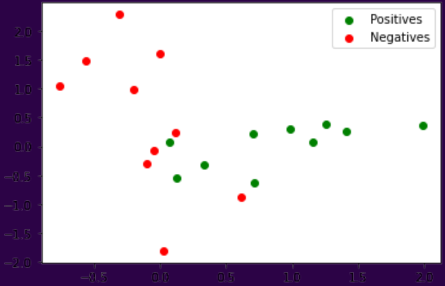


Figure: https://www.pinterest.com/pin/672232681855858622/?lp=true

Classification

Given the dataset (x_i, y_i) for $i = 1, 2, \ldots, N$, find a function $f(x)$ (model) so that it can predict the label $\hat{y}$ for some input x, even if it is not in the dataset, i.e. $\hat{y} = f(x)$.

- Positive : $y = 1$
- Negative : $y = 0$



Classification via regression

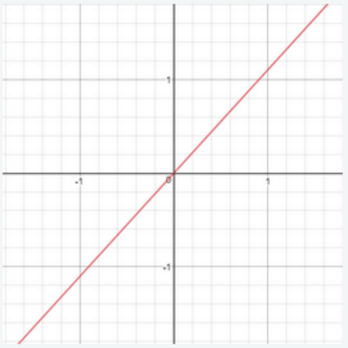
- Proposal: train a model to fit the data with linear regression!

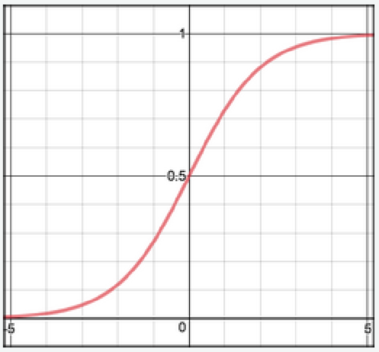
Classification via regression

- Proposal: train a model to fit the data with linear regression!
- What could be the problem?

Sigmoid Function

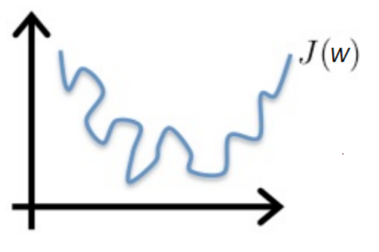
- Recall from linear regression $z = w_0 + w_1 x$
- By applying the sigmoid function to z, we enforce $0 \leq \hat{y} \leq 1$
  - $\hat{y} = sigmoid(z) = \frac{1}{1+e^{-z}}$

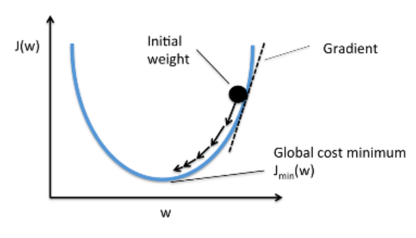




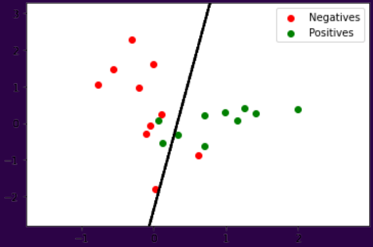
Classification Loss Function

- Cannot use the same cost function that we used for linear regression
  - MSE of a logistic function has many local minima
- Use $\frac{1}{N}\sum_{i=1}^{N}\left[-ylog(\hat{y})-(1-y)log(1-\hat{y})\right]$
  - This loss function is called binary cross entropy loss
  - This loss function has only one minimum





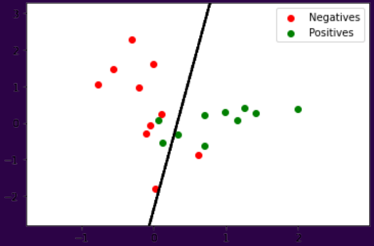
Decision Boundary



- Evaluation metric :

$$\text{Accuracy} = \frac{\text{Number of correct prediction}}{\text{Total number of prediction}}$$

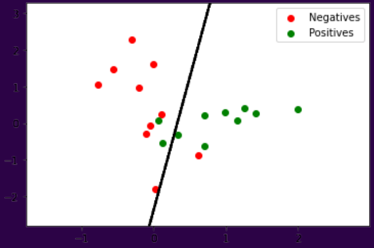
Decision Boundary



- Evaluation metric :

$$\text{Accuracy} = \frac{\text{Number of correct prediction}}{\text{Total number of prediction}}$$

- What is the accuracy in this example ?



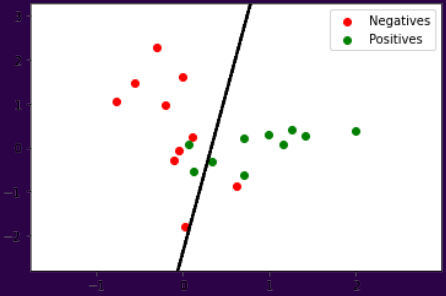
- Evaluation metric :

$$\text{Accuracy} = \frac{\text{Number of correct prediction}}{\text{Total number of prediction}} = \frac{17}{20} = 0.85 = 85\%$$

Types of Errors in Classification

- Correct predictions:
  - True Positive (TP) : Predict $\hat{y} = 1$ when $y = 1$
  - True Negative (TN) : Predict $\hat{y} = 0$ when $y = 0$
- Two types of errors:
  - False Positive/ False Alarm (FP): $\hat{y} = 1$ when $y = 0$
  - False Negative/ Missed Detection (FN): $\hat{y} = 0$ when $y = 1$

Example



- How many True Positive (TP) are there ?
- How many True Negative (TN) are there ?
- How many False Positive (FP) are there ?
- How many False Negative (FN) are there ?

Performance metrics for a classifier

- Accuracy of a classifier:
  - (TP + TF)/(TP+FP+TN+FN) (percentage of correct classification)
- Why accuracy alone is not a good measure for assessing the model

Performance metrics for a classifier

- Accuracy of a classifier:
  - (TP + TF)/(TP+FP+TN+FN) (percentage of correct classification)
- Why accuracy alone is not a good measure for assessing the model
  - There might be an overwhelming proportion of one class over another (unbalanced classes)
  - Example: A rare disease occurs 1 in ten thousand people
  - A test that classifies everyone as free of the disease can achieve 99.999% accuracy when tested with people drawn randomly from the entire population

Other metrics

Some other metrics

- Sensitivity/Recall/TPR = TP/(TP+FN) (How many positives are detected among all positive?)
- Precision = TP/(TP+FP) (How many detected positives are actually positive?)
- Specificity/TNR = TN/(TN+FP) (How many negatives are detected among all negatives?)

Exercise: think of tasks for which sensitivity, precision, or specificity is a better metric.

Outline

Demo: Diagnosing Breast Cancer

- We're going to use the breast cancer dataset to predict whether the patients' scans show a malignant tumour or a benign tumour.
- Let's try to find the best linear classifier using logistic regression.

Outline

Multiclass Classificaiton

- Previous model: $f(\mathbf{x}) = \sigma(\mathbf{w}^T \phi(\mathbf{x}))$
- Representing Multiple Classes:
  - One-hot / 1-of-K vectors, ex : 4 Class
  - Class 1 : $\mathbf{y} = [1, 0, 0, 0]$
  - Class 2 : $\mathbf{y} = [0, 1, 0, 0]$
  - Class 3 : $\mathbf{y} = [0, 0, 1, 0]$
  - Class 4 : $\mathbf{y} = [0, 0, 0, 1]$

Multiclass Classificaiton

- Previous model: $f(\mathbf{x}) = \sigma(\mathbf{w}^T \phi(\mathbf{x}))$
- Representing Multiple Classes:
  - One-hot / 1-of-K vectors, ex : 4 Class
  - Class 1 : $\mathbf{y} = [1, 0, 0, 0]$
  - Class 2 : $\mathbf{y} = [0, 1, 0, 0]$
  - Class 3 : $\mathbf{y} = [0, 0, 1, 0]$
  - Class 4 : $\mathbf{y} = [0, 0, 0, 1]$
- Multiple outputs: $f(\mathbf{x}) = \text{softmax}(W^T \phi(\mathbf{x}))$
- Shape of $W^T \phi(\mathbf{x})$: $(K, 1) = (K, D) \times (D, 1)$
- $\text{softmax}(\mathbf{z})_k = \frac{e^{z_k}}{\sum_j e^{z_j}}$

Multiclass Classificaiton

- Multiple outputs: $f(\mathbf{x}) = \text{softmax}(\mathbf{z})$ with $\mathbf{z} = W^T\phi(\mathbf{x})$
- $\text{softmax}(\mathbf{z})_k = \frac{e^{\mathbf{z}_k}}{\sum_j e^{\mathbf{z}_j}}$
- Softmax example: If $\mathbf{z} = \begin{bmatrix} -1 \\ 2 \\ 1 \\ -4 \end{bmatrix}$ then,

$$\text{softmax}(z) = \begin{bmatrix} \frac{e^{-1}}{e^{-1}+e^{2}+e^{1}+e^{-4}} \\ \frac{e^{2}}{e^{-1}+e^{2}+e^{1}+e^{-4}} \\ \frac{e^{1}}{e^{-1}+e^{2}+e^{1}+e^{-4}} \\ \frac{e^{-4}}{e^{-1}+e^{2}+e^{1}+e^{-4}} \end{bmatrix} \approx \begin{bmatrix} 0.035 \\ 0.704 \\ 0.259 \\ 0.002 \end{bmatrix}$$

Cross-entropy

- Multiple outputs: $\hat{\mathbf{y}}_{\mathbf{i}} = \text{softmax}(W^T \phi(\mathbf{x}_i))$
- Cross-Entropy: $J(W) = -\sum_{i=1}^{N} \sum_{k=1}^{K} \mathbf{y}_{ik} log(\hat{\mathbf{y}}_{ik})$
- Example : K = 4

$$\text{If, } \mathbf{y}_i = [0, 0, 1, 0] \text{ then, } \sum_{k=1}^{K} \mathbf{y}_{ik} log(\hat{\mathbf{y}}_{ik}) = log(\hat{\mathbf{y}}_{i3})$$

Outline

Mini Project

- Task: Predict fish weight!
- You should split the given dataset into training and validation set.
- Test set will be released on Sunday night.
- Next Monday morning: present your project and the model performance on the test set.
- Each team should present for 8-10 minutes.

Presentation Template

- Slide 1: Title and introduction
- Slide 2: Your model and loss function
- Slide 3 & 4: What is your choice of feature transformation, regularizer (Ridge/Lasso?) hyper-parameters, etc.
- Slide 5: Model performance on training and test set?
- Slide 6: Challenges and how you resolve them.
- Slide 7: Conclusion

Thank You!

- Next Week: Deep Learning
- Have a fun weekend!
- Revise Revise Revise!