Name: Date: Class:

Classification

You’re managing an advertising budget for a snack company. Each ad impression costs money, so you want to focus on users who are most likely to click on the ice cream ad. You have historical data about 12 users and whether they clicked.

User	Device	Time	Demographic	Location	Clicked
1	Mobile	Evening	Teen	School	Yes
2	Desktop	Morning	Parent	Work	No
3	Tablet	Afternoon	Young Adult	Home	Yes
4	Desktop	Night	Senior	Home	No
5	Tablet	Morning	Young Adult	Public	No
6	Desktop	Morning	Parent	Work	Yes
7	Mobile	Afternoon	Teen	School	No
8	Tablet	Night	Senior	Public	Yes
9	Mobile	Morning	Parent	Home	Yes
10	Mobile	Evening	Teen	Public	No
11	Tablet	Evening	Senior	Public	No
12	Desktop	Afternoon	Teen	Home	Yes

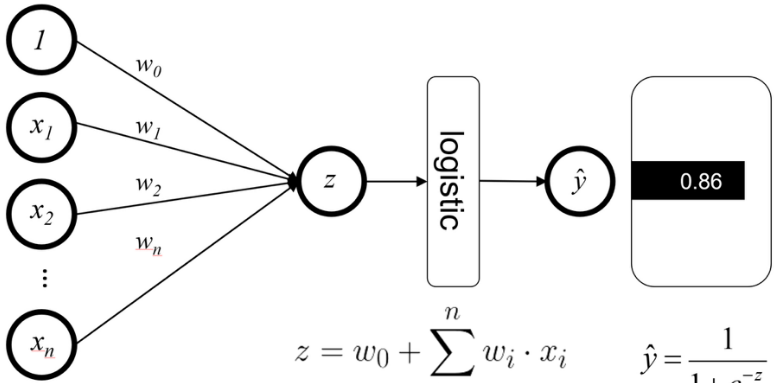
Based on this historical data, can you predict whether these new users will click on your ad?

New User	Device	Time	Demographic	Location
A	Mobile	Evening	Young Adult	Home
B	Desktop	Morning	Senior	Work
C	Tablet	Afternoon	Teen	Public
D	Mobile	Night	Parent	School
E	Desktop	Evening	Young Adult	Public

In this exercise, you’ll practice applying five different classifiers to the same dataset. You’ll compute predictions step by step to understand how each model works and compare their decisions.

Name: Date: Class:

Logistic Regression



$$z = w_0 + \sum_{i=1}^{n} w_i \cdot x_i \qquad \hat{y} = \frac{1}{1+e^{-z}}$$

Feature Weights:

Device:

Time:

Demographic:

Location:

Bias:

Example:

Predict whether New User B will click on the ad using logistic regression:

B	Desktop	Morning	Senior	Work

Name: Date: Class:

<u>Naïve Bayes</u>

$$\underbrace{P(c \mid x)}_{\text{Posterior Probability}} = \frac{\overbrace{P(x \mid c)}^{\text{Likelihood}}\,\overbrace{P(c)}^{\text{Class Prior Probability}}}{\underbrace{P(x)}_{\text{Predictor Prior Probability}}}$$

$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

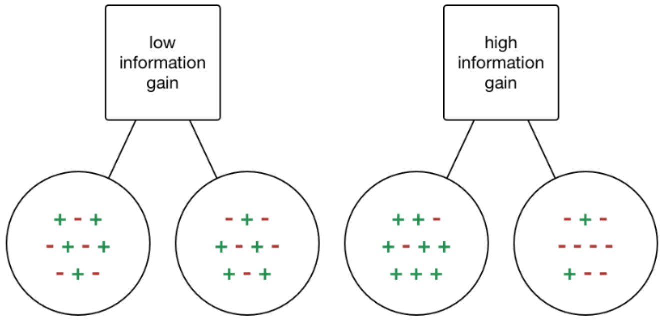
Example:

Predict whether New User C will click on the ad using naïve bayes classification:

C	Tablet	Afternoon	Teen	Public

Name: Date: Class:

Decision Trees



$$Entropy(S) = -\sum_{c \in C} P(S_c) \log_2 P(S_c)$$

$$Information\ Gain(S, X) = E(S) - \sum_{v \in Values(X)} \frac{|S_{X=v}|}{|S|} \cdot E(S_{X=v})$$

Example:

Decision Tree:

Use the decision tree to predict if New User D will click on the ad:

D Mobile Night Parent School

Name: Date: Class:

Random Forest

Pick one feature at random.

Pick five users from the dataset. List their IDs.

Construct a decision tree of those five users using the chosen feature. Draw the decision tree here.

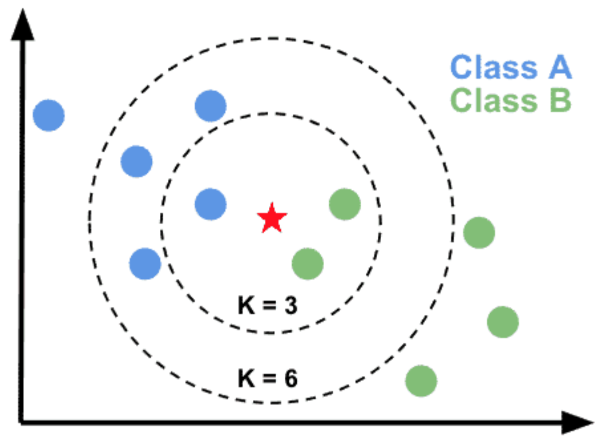
Use the new decision tree to predict if New User D will click on the ad:

D	Mobile	Night	Parent	School

How did the rest of the decision trees vote on the classification of this user?

Name: Date: Class:

K-Nearest Neighbors



Calculate the *Hamming distance* between each datapoint and New User E.

E Desktop Evening Young Adult Public

User	Device	Time	Demographic	Location	Clicked	Distance
1	Mobile	Evening	Teen	School	Yes	
2	Desktop	Morning	Parent	Work	No	
3	Tablet	Afternoon	Young Adult	Home	Yes	
4	Desktop	Night	Senior	Home	No	
5	Tablet	Morning	Young Adult	Public	No	
6	Desktop	Morning	Parent	Work	Yes	
7	Mobile	Afternoon	Teen	School	No	
8	Tablet	Night	Senior	Public	Yes	
9	Mobile	Morning	Parent	Home	Yes	
10	Mobile	Evening	Teen	Public	No	
11	Tablet	Evening	Senior	Public	No	
12	Desktop	Afternoon	Teen	Home	Yes	

Who are the 3 closest users to New User E? Do they think that E will click on the ad?

Name: Date: Class:

Comparison

Visit sklearn-classifiers-playground. For each dataset, rank the classifiers according to performance.

	Concentric Circles	Spiral	Moons	Quadrants	Random
Logistic Regression					
2-Layer Neural Network					
Decision Tree					
Random Forest					
K-nearest neighbors					

Which classifier would you use if:

- You had a dataset with a lot of noise?
- You had a huge, high-dimensional dataset?
- You wanted to know which features are most important?
- You need it to be trained fast?
- You need a very fast prediction for new data?

Explain why.