Recitation 10
SVMs, Kernels, Naïve Bayes

10-301/10-601: Introduction to Machine Learning

12/04/2020

1 Naive Bayes

Definitions:

- Bayes Rule: $p(A|B) = \frac{p(A,B)}{p(B)} = \frac{p(B|A)p(A)}{p(B)}$
- Naive Bayes Model:

  assumes attributes $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_K$ are conditionally independent of each other given the class label y, and that their prediction can be written as $\hat{y} = \arg\max_y P(y|X)$, where

$$p(y|X = (\mathbf{x}_1, ..., \mathbf{x}_k)) \propto p(y)p(X|y) = p(y)\prod_{k=1}^{K} p(\mathbf{x}_k|y)$$

Questions:

1. What are some differences and similarities between Generative Models and Discriminative Classifiers?

2. You love your old red mustang convertible, but you just couldn't keep up on the payments. It's now time to sell the car. To generate some interest, you decide to post an online advertisement for your car. When someone views the advertisement, you know three things - are they currently looking for a car? Do they like the color red? Do they currently have free time to look at an ad? The below table summarizes some previous data you collected. Note 1 represents "Yes" and 0 represents "No".

Clicked?	Looking?	Likes Red?	Free?
1	1	1	1
0	0	1	0
0	0	1	1
1	0	0	0
0	0	1	1
0	0	1	1
1	1	1	1
1	1	0	1
0	0	0	0

A new person visits the page. What you know is that they are looking for a car, don't like red, and currently have free time. Using a Naive Bayes approach, Do you predict that they will click on the advertisement? A table with parameter values is provided below for convenience.

feature	Clicked = 1	Clicked = 0
Looking?	3/4	0/5
Likes Red?	2/4	4/5
Free?	3/4	3/5

3. Is there anything wrong here? What do we know will happen to our prediction for any person if looking = 1? Why might this be a problem? How can we fix it?

4. **BONUS** When Y is Boolean and $\mathbf{X} = (X_1, \cdots, X_n)$ is a vector of continuous variables, then the assumptions of the Gaussian Naive Bayes classifier imply that $P(Y \mid \mathbf{X})$ is given by the logistic function with appropriate parameters W. In particular:

$$P(Y = 1 \mid \mathbf{X}) = \frac{1}{1 + \exp(b + \sum_{i=1}^{n} w_i X_i)}$$

and

$$P(Y = 0 \mid \mathbf{X}) = \frac{\exp(b + \sum_{i=1}^{n} w_i X_i)}{1 + \exp(b + \sum_{i=1}^{n} w_i X_i)}$$

Consider instead the case where Y is Boolean and $\mathbf{X} = (X_1, \cdots, X_n)$ is a vector of Boolean variables. Prove for this case also that $P(Y \mid \mathbf{X})$ follows this same form (and hence that Logistic Regression is also the discriminative counterpart to a Naive Bayes generative classifier over Boolean features).

Hints

1. Simple notation will help. Since the X_i are Boolean variables, you need only one parameter to define $P(X_i \mid Y = y_k)$. Define $\phi_{i1} \equiv P(X_i = 1 \mid Y = 1)$, in which case $P(X_i = 0 \mid Y = 1) = (1 - \phi_{i1})$. Similarly, use ϕ_{i0} to denote $P(X_i = 1 | Y = 0)$.

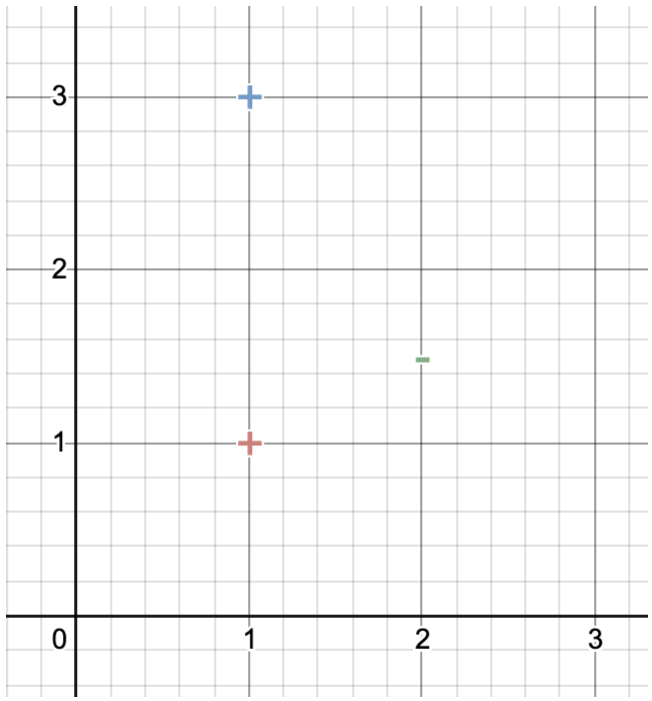
2. Notice with the above notation you can represent $P(X_i \mid Y = 1)$ as follows

$$P(X_i \mid Y = 1) = \phi_{i1}^{(X_i)} (1 - \phi_{i1})^{(1 - X_i)}$$

Note when $X_i = 1$ the second term is equal to 1 because its exponent is zero. Similarly, when $X_i = 0$ the first term is equal to 1 because its exponent is zero.

2 SVMs

1. What is the decision boundary and the margin if we run a Hard-Margin SVM on the following set of points?



2. A few additional data points are added to the data set in figures 2 (a) and 2 (b). Draw the new decision boundaries and give the margins corresponding to this boundaries. In which case does the decision boundary undergo a change and why?

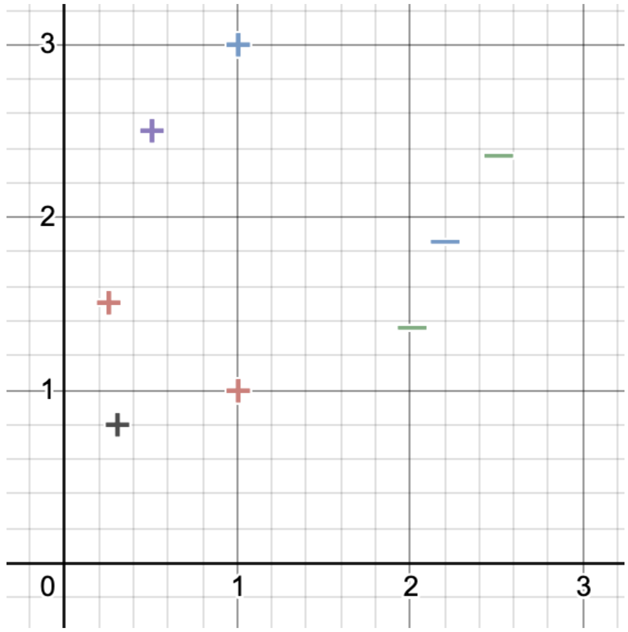


Figure 2(a)

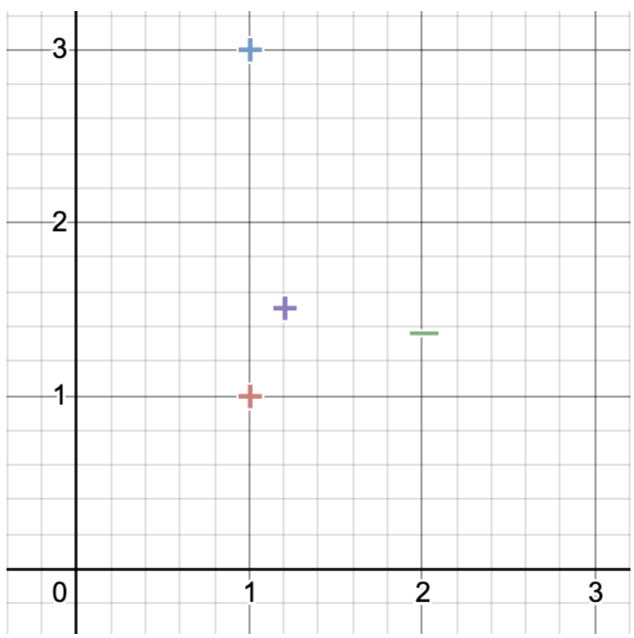
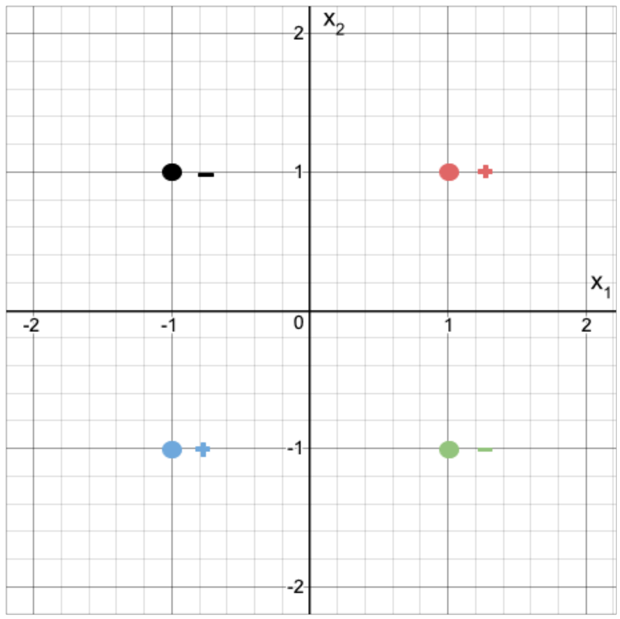


Figure 2(b)

3 Kernels

The XOR-problem is a non-linear problem which can be represented by the plot below.



For the following questions, consider a feature transformation - $\phi([x_1, x_2]^T) = [x_1, x_1x_2]^T$

1. What is the kernel $K(x, z)$?
2. How is the dataset represented in the transformed space?
3. Is the dataset linearly separable in the transformed space? If so, give the boundary in the original space?