CSE 250B: Section 2 - Sharad Vikram

1. Decision Tree Practice (borrowed from Berkeley)

   For each example on the back of this page,

   1. Draw the decision tree of depth at most 2 that can separate the given data (pluses and dots) completely, by filling in binary predicates (which only involve thresholding of a *single* variable) in the boxes for the decision trees below. If the data is already separated when you hit a box, simply write the class, and leave the sub-tree hanging from that box empty.
   2. Draw the corresponding decision boundaries on the scatter plot, and write the class labels for each of the resulting bins somewhere inside the resulting bins.

   If the data can not be separated completely by a depth 2 decision tree, simply cross out the tree template. We solve the first part as an example.

2. DNAive Bayes

   Given a DNA strand X, let's try to predict whether or not the user has a disease ($Y = 1$) or not ($Y = 0$) using Naive Bayes. (Disclaimer: this example has no basis in real biology.)

```
1. g g g g g g a g c g (disease)
2. a a a g a g a a g a (no disease)
3. c g a g c c g g c c (disease)
4. a a g t a g g c a a (no disease)
5. a a c c g a c g a a (no disease)
6. t a a a g a t a a a (no disease)
7. c g g g t a g g c a (disease)
8. a a t a c g c a g g (no disease)
9. a c a a a a a g c g (no disease)
10.t a a a g a a g a a (no disease)
```

   Remember that we assume that *order does not matter.* Therefore, for a strand x,

$$P(Y = y|X = x) \propto P(X = x|Y = y)P(Y = y) = P(Y = y)\prod_{i=1}^{N} P(x_i|Y = y)$$

   1. Estimate the prior distribution, $P(Y = y)$.

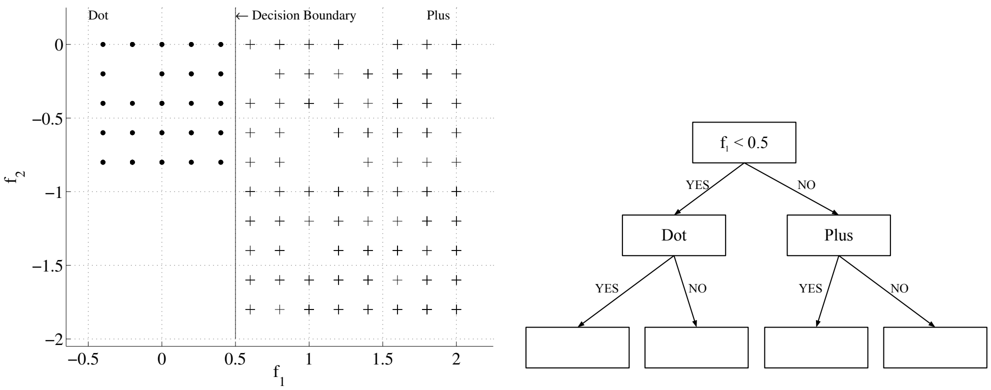
$Y = 0$	
$Y = 1$	

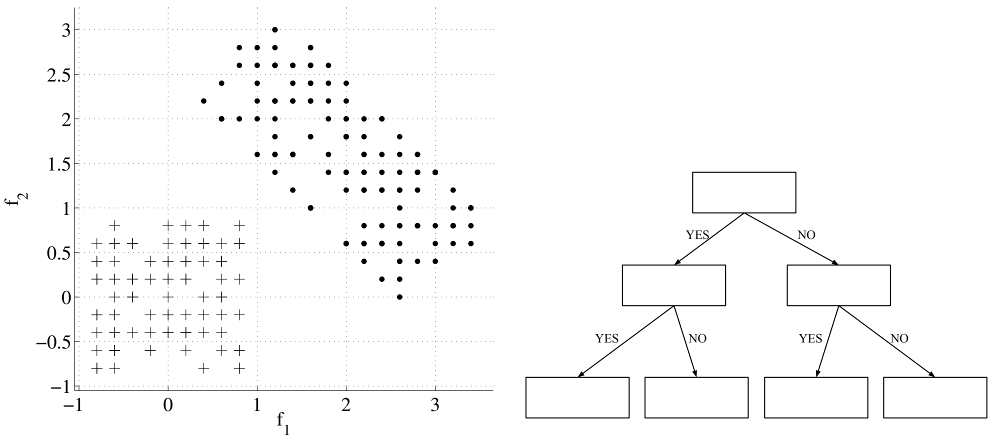
   2. Estimate the class conditional distribution, $P(x_i|Y = y)$ with smoothing ($\alpha = 1$).

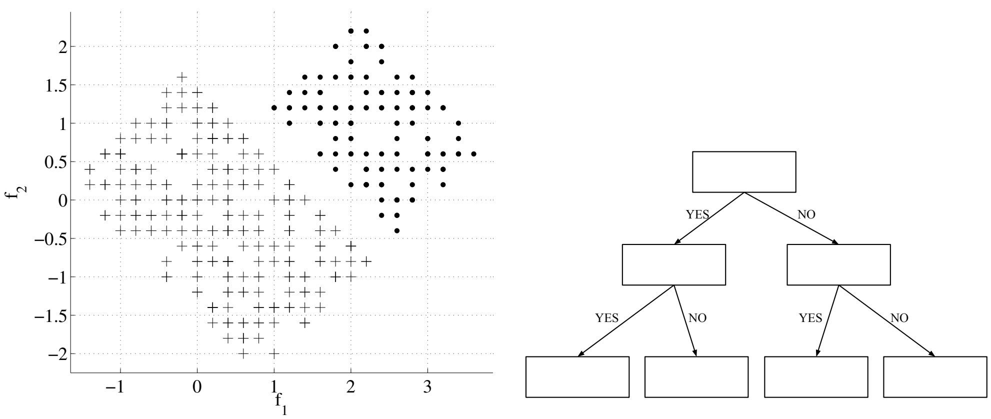
	a	c	g	t
$Y = 0$				
$Y = 1$				

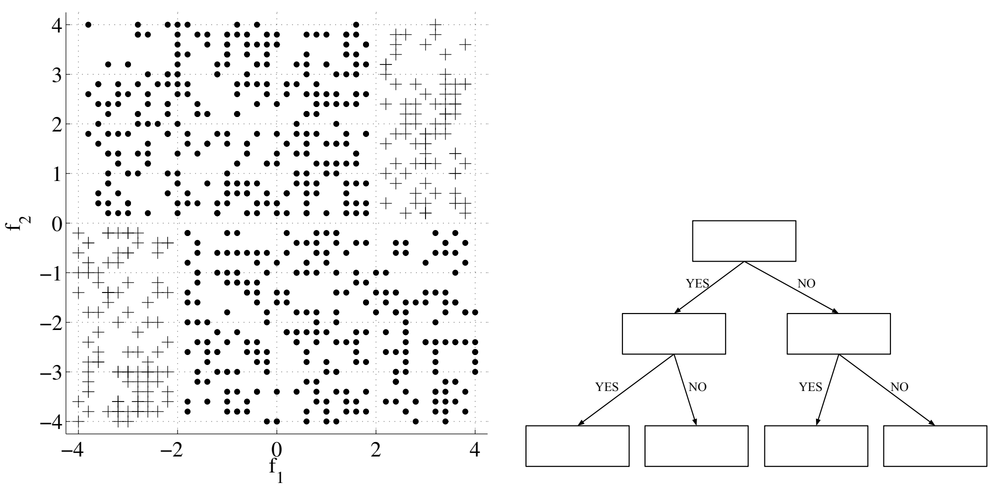
   3. Predict the class of the following examples using the prediction rule $h(x) = \text{argmax}_y P(Y = y|X = x)$.

```
1. g g g a g g a a a a
2. c t g t g c g g g c
```


Dot
← Decision Boundary
Plus
f_2
f_1
$f_1 < 0.5$
YES
NO
Dot
Plus
YES
NO
YES
NO


f_2
f_1
YES
NO
YES
NO
YES
NO


f_2
f_1
YES
NO
YES
NO
YES
NO


f_2
f_1
YES
NO
YES
NO
YES
NO