CS 5350/6350: Machine Learining Spring 2015

Homework 1 Solution

Handed out: Jan 21, 2015
Due date: Feb 4, 2015

1 Decision Trees

1.1 Boolean Functions

In some cases there may be more than one equivalent tree.

1. $(A \vee B) \wedge C$ [3 points]

```
if C:
    if A:
         class True
     if not A:
         if B:
             class True
         if not B:
             class False
if not C:
    class False
```

2. $A \oplus B$ [3 points]

```
if not A:
    if not B:
        class False
    if B:
         class True
if A:
    if not B:
        class True
    if B:
        class False
```

3. $A \wedge \neg B \wedge \neg C \wedge D$ [3 points]

```
if A:
    if not B:
        if not C:
            if D:
                class True
            if not D:
                class False
        if C:
            class False
    if B:
        class False
if not A:
    class False
```

1.2 Building Decision Trees

$$H(S) = -\sum_{k \in S} P(k) \log_2 P(k)$$

$$ME(S) = 1 - \max_{k \in S} P(k)$$

$$IG(A, S) = H(S) - \sum_{v \in S(A)} P(v) H(v)$$

(a) Find the entropy of the Balloon dataset. [2 points]

$$H(Balloon) = -\frac{12}{20} \log_2 \left(\frac{12}{20}\right) - \frac{8}{20} \log_2 \left(\frac{8}{20}\right) = 0.97$$

(b) Find the information gain of the Action feature. [2 points]

$$IG(Action, Balloon) = 0.97 - \frac{8}{20} H(Stretch) - \frac{12}{20} H(Dip)$$

$$IG(Action, Balloon) = 0.97 - \frac{8}{20}(0) - \frac{12}{20}(0.92) = 0.42$$

(c) Use the ID3 heuristic we have seen in the class to manually construct a decision tree that correctly classifies the data. [7 points]

Find the information gain for each feature.

Feature	Information Gain
Color	0.0
Size	0.0
Action	0.42
Age	0.42

Choose the action feature. The two values of Action are STRETCH and DIP. For STRETCH all the labels are True, so make a leaf. Find the information gains for the subset of data for which Action is DIP.

Feature	Information Gain
Color	0.0
Size	0.0
Age	0.92

Chose the age feature. The two sub-trees both classify the remaining data perfectly, so make two leaf nodes. The final tree is

```
Action
 +--STRETCH---True
 +--DIP---Age
           +--CHILD---False
           +--ADULT---True
```

(d) What is the information gain of the Action feature using $MajorityError$ as a measure of impurity? [3 points]

$$IG(Action, Balloon) = 0.4 - \frac{8}{20}ME(Stretch) - \frac{12}{20}ME(Dip)$$

$$IG(Action, Balloon) = 0.4 - \frac{8}{20}(0) - \frac{12}{20}(.33) = 0.2$$

(e) Using the $MajorityError$-based impurity measure, construct a decision tree for the balloons data. [7 points]

Find the information gain for each feature.

Feature	Information Gain
Color	0.0
Size	0.0
Action	0.2
Age	0.2

Choose the action feature. The two values of Action are STRETCH and DIP. For STRETCH all the labels are True, so make a leaf. Find the information gains for the subset of data for which Action is DIP.

Feature	Information Gain
Color	0.0
Size	0.0
Age	0.33

Chose the age feature. The two sub-trees both classify the remaining data perfectly, so make two leaf nodes. The final tree is

```
Action
 +--STRETCH---True
 +--DIP---Age
          +--CHILD---False
          +--ADULT---True
```

1.3 Train and Test

The procedure and resulting trees are the same as in 1.2.c and 1.2.e. The information gains are shown in the tables below.

For the whole data set.

Feature	Entropy IG	Majority Error IG
Color	0.11	0.0
Size	0.01	0.0
Action	0.50	0.17
Age	0.34	0.08

For the Action=DIP subset.

Feature	Entropy IG	Majority Error IG
Color	0.0	0.0
Size	0.0	0.0
Age	0.92	0.33

2 Nearest Neighbors

2.1 Euclidean Distance Map

See Figure 1

2.2 Manhattan Distance Map

See Figure 2

2.3 Weighted Nearest Neighbor

A: $\frac{2}{\sqrt{1^2+3^2}} = 0.63$

B: $\frac{1}{\sqrt{1^2+5^2}} = 0.20$

C: $\frac{1}{\sqrt{2^2+2^2}} = 0.35$

3 The Badges Game (again)

The table below shows the accuracy of ID3 and KNN using two features, the first and last letter of the last name. The table includes two types of nearest neighbor classifiers. One uses the euclidean distance, treating characters as numbers, and the other uses the hamming distance between the characters. Since the data is close to linearly separable using

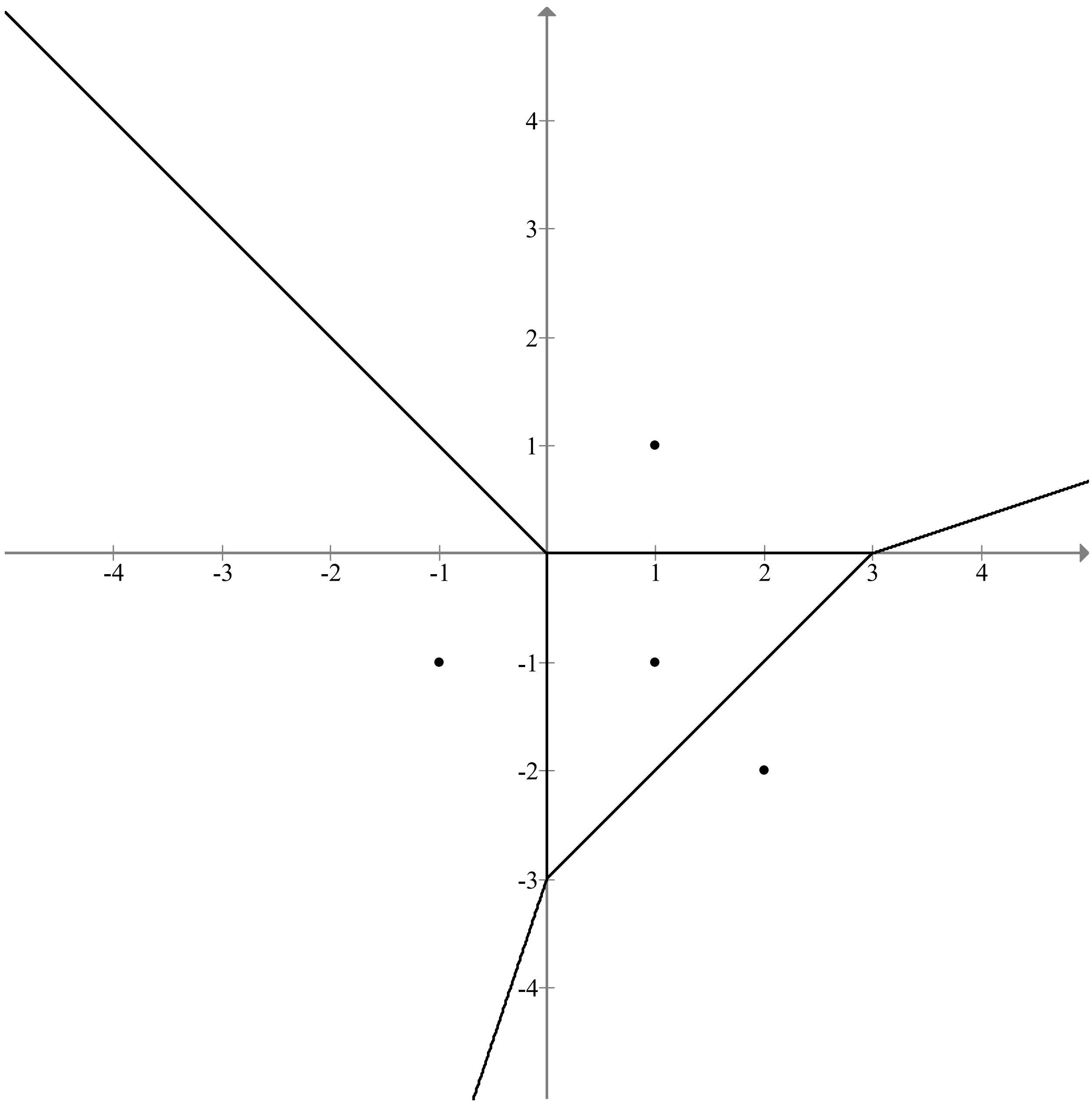

Figure 1: The Voronoi map for the L2 norm.

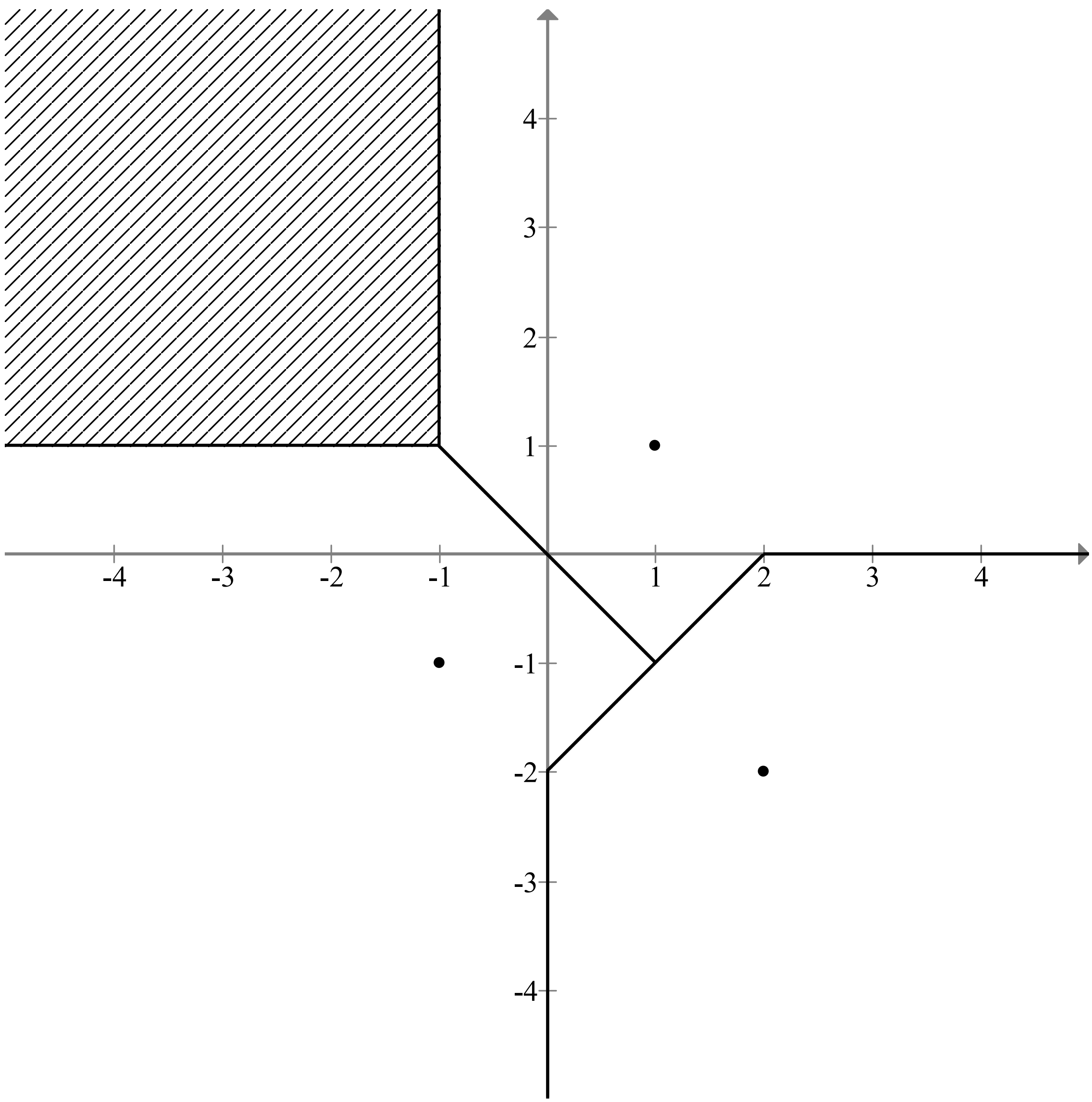

Figure 2: The Voronoi map for the L1 norm.

these features, euclidean KNN approximates the line well, where the hamming distance and decision tree need more data to get a good representation of the line between the classes.

Algorithm	Accuracy
MajorityError	0.82
Entropy	0.82
1NN	0.96
2NN	0.96
3NN	0.95
4NN	0.96
5NN	0.94
1NN Hamming	0.83
2NN Hamming	0.83
3NN Hamming	0.76
4NN Hamming	0.82
5NN Hamming	0.79