CS 343: Artificial Intelligence
Perceptrons

Prof. Yuke Zhu — The University of Texas at Austin

[These slides based on those of Dan Klein and Pieter Abbeel for CS188 Intro to AI at UC Berkeley. All CS188 materials are available at http://ai.berkeley.edu.]
Error-Driven Classification
Errors, and What to Do

- Examples of errors

Dear GlobalSCAPE Customer,

GlobalSCAPE has partnered with ScanSoft to offer you the latest version of OmniPage Pro, for just $99.99* - the regular list price is $499! The most common question we've received about this offer is - Is this genuine? We would like to assure you that this offer is authorized by ScanSoft, is genuine and valid. You can get the . . .

. . . To receive your $30 Amazon.com promotional certificate, click through to

http://www.amazon.com/apparel

and see the prominent link for the $30 offer. All details are there. We hope you enjoyed receiving this message. However, if you'd rather not receive future e-mails announcing new store launches, please click . . .
What to Do About Errors

- Problem: there’s still spam in your inbox

- Need more features – words aren’t enough!
  - Have you emailed the sender before?
  - Have 1M other people just gotten the same email?
  - Is the sending information consistent?
  - Is the email in ALL CAPS?
  - Do inline URLs point where they say they point?
  - Does the email address you by (your) name?

- Naïve Bayes models can incorporate a variety of features, but tend to do best when homogeneous (e.g. all features are word occurrences) and/or roughly independent
Linear Classifiers
Hello,
Do you want free printr cartriges? Why pay more when you can get them ABSOLUTELY FREE! Just

```
# free          : 2
YOUR_NAME     : 0
MISSPELLED   : 2
FROM_FRIEND  : 0
...

PIXEL-7,12    : 1
PIXEL-7,13    : 0
...
NUM_LOOP      : 1
...
```

```
SPAM or +
```

```
"2"
```
Some (Simplified) Biology

- Very loose inspiration: human neurons
Linear Classifiers

- Inputs are feature values
- Each feature has a weight
- Sum is the activation

\[
\text{activation}_w(x) = \sum_i w_i \cdot f_i(x) = w \cdot f(x)
\]

- If the activation is:
  - Positive, output +1
  - Negative, output -1
Weights

- Binary case: compare features to a weight vector
- Learning: figure out the weight vector from examples

\[
\begin{align*}
# \text{free} & : 4 \\
\text{YOUR\_NAME} & : -1 \\
\text{MISSPELLED} & : 1 \\
\text{FROM\_FRIEND} & : -3 \\
\ldots & \\
\end{align*}
\]

\[
\begin{align*}
# \text{free} & : 2 \\
\text{YOUR\_NAME} & : 0 \\
\text{MISSPELLED} & : 2 \\
\text{FROM\_FRIEND} & : 0 \\
\ldots & \\
\end{align*}
\]

\[
\begin{align*}
# \text{free} & : 0 \\
\text{YOUR\_NAME} & : 1 \\
\text{MISSPELLED} & : 1 \\
\text{FROM\_FRIEND} & : 1 \\
\ldots & \\
\end{align*}
\]

Dot product $w \cdot f$ positive means the positive class
Decision Rules
Binary Decision Rule

- In the space of feature vectors
  - Examples are points
  - Any weight vector is a hyperplane
  - One side corresponds to $Y=+1$
  - Other corresponds to $Y=-1$

$$w$$

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIAS</td>
<td>-3</td>
</tr>
<tr>
<td>free</td>
<td>4</td>
</tr>
<tr>
<td>money</td>
<td>2</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

$$f \cdot w = 0$$

+1 = SPAM
-1 = HAM
Weight Updates
Learning: Binary Perceptron

- Start with weights = 0
- For each training instance:
  - Classify with current weights
    - If correct (i.e., \( y = y^* \)), no change!
    - If wrong: adjust the weight vector
Learning: Binary Perceptron

- Start with weights = 0
- For each training instance:
  - Classify with current weights
    $$y = \begin{cases} 
    +1 & \text{if } w \cdot f(x) \geq 0 \\
    -1 & \text{if } w \cdot f(x) < 0 
    \end{cases}$$
  - If correct (i.e., $y=y^*$), no change!
  - If wrong: adjust the weight vector by adding or subtracting the feature vector. Subtract if $y^*$ is -1.
    $$w = w + y^* \cdot f$$
Examples: Perceptron

- Separable Case
Multiclass Decision Rule

- If we have multiple classes:
  - A weight vector for each class:
    \[ w_y \]
  - Score (activation) of a class \( y \):
    \[ w_y \cdot f(x) \]
  - Prediction highest score wins
    \[ y = \arg \max_y w_y \cdot f(x) \]

Binary = multiclass where the negative class has weight zero
Learning: Multiclass Perceptron

- Start with all weights = 0
- Pick up training examples one by one
- Predict with current weights
  \[ y = \arg \max_y \ w_y \cdot f(x) \]
- If correct, no change!
- If wrong: lower score of wrong answer, raise score of right answer
  \[ w_y = w_y - f(x) \]
  \[ w_{y^*} = w_{y^*} + f(x) \]
Properties of Perceptrons

- Separability: true if some parameters get the training set perfectly correct

- Convergence: if the training is separable, perceptron will eventually converge (binary case)

- Mistake Bound: the maximum number of mistakes (binary case) related to the margin or degree of separability

\[
\text{mistakes} < \frac{k}{\delta^2}
\]
Examples: Perceptron

- Non-Separable Case
Perceptron Exercises

For each of the datasets represented by the graphs below, please select the feature maps for which the perceptron algorithm can perfectly classify the data.

Each data point is in the form \((x_1, x_2)\), and has some label \(Y\), which is either a 1 (dot) or −1 (cross).