Perceptron

Linear Threshold Algorithm
Perceptron

\[ a = \sum_j w_i x_j \]

output = \( f(a_i) \)

Can also have bias \( b \) (fixed to 1)

input values, \( x_i \)'s
Each \( x_i \) weighted by \( w_i \)

\[ f(a) = \frac{1}{1+\exp(-a)} \]

f(a) step function

\[ f(a) = +1 \]

\[ f(a) = -1 \]
Classifying points – illustration of LTU “wiggle room”

From Duda and Hart, 1973
Perceptron Algorithm

• Keeps weights $w_j$, one per feature
• Online algorithm, initially $w = (0,\ldots,0)$
• Repeat (until consistent with data):
  get next training example $i$: $(x_i,t_i)$
  if $(w \cdot x_i) t_i \leq 0$ then mistake:
    $w$ gets $w + \eta_i t_i x_i$

$\eta_i$ values are learning rates (step sizes)
Perceptron Class Exercise:

- Assume $\eta_i$ always 1

<table>
<thead>
<tr>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>+1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>-1</td>
</tr>
<tr>
<td>-3</td>
<td>1</td>
<td>+1</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>

(gap $\approx 2/15$)
Perceptron Convergence

• If data linearly separable with “gap” then converges within \((1/\text{gap})^2\) mistakes when instances normalized to length 1.

• For data not linearly separable it converges if (Robbins-Monro alg.):
  \[ \eta_i \text{ values go to 0 (as } i \text{ goes to } \infty) \]
  \[ \text{sum of } \eta_i \text{ values goes to } \infty \]
  \[ \text{sum of } (\eta_i)^2 \text{ values finite} \]
Perceptron as gradient descent

- **Perceptron criteria**: minimize “badness” of mistake on example $i$:
  $$-t_i (w \cdot x_i)$$

- Differentiate wrt $w_j$ gives gradient component:
  $$-t_i x_{i,j}$$

- Negative gradient, $t_i x_i$ is direction of steepest descent, add $t_i x_{i,j}$ to $w_j$ (for each $j$) or equivalently add vector $t_i x_i$ to $w$
Perceptron notes

• Can run in batch mode - save updates until completed pass through data
• Voted perceptron idea
• Multiclass- learn a $w_y$ for each class, predict with $y$ maximizing $w_y \cdot x$
• Learns classifier directly (no probability)