IAML: The Perceptron

Nigel Goddard and Victor Lavrenko
School of Informatics

Semester 1

A Simple Linear Algorithm

- Can we do something simpler than logistic regression? And still be linear?
- For logistic regression we had this squashing function

$$f(z) = \sigma(z) \equiv 1/(1 + \exp(-z))$$

- What if we just have a step function?

$$f(z) = \text{sign}[z] = \begin{cases} 1 & \text{if } z \geq 0 \\ -1 & \text{otherwise} \end{cases}$$

- Notice that we call the classes $y \in \{-1, 1\}$. This is just for convenience later on.
- This architecture is called a *perceptron*, and has a very long history.

Classifying Using a Perceptron

- Like any other linear classifier. Given $\tilde{\mathbf{w}}$, w_0 and a $\mathbf{x}$ to classify, do
$$\hat{y} = \begin{cases} 1 & \text{if } \tilde{\mathbf{w}}^T\mathbf{x} + w_0 \geq 0 \\ -1 & \text{otherwise} \end{cases}$$
- This is OK, but how are you going to train it?
- The problem is that you can't use gradient descent anymore.

The Perceptron Learning Rule

- The following rule was studied by Rosenblatt (1956)

  repeat
      for i in $1, 2, \ldots n$
          $\hat{y} \leftarrow \text{sign}[\mathbf{w}^T \mathbf{x}_i]$
          if $\hat{y} \neq y_i$
              $\mathbf{w} \leftarrow \mathbf{w} + y_i \mathbf{x}_i$
  until all training examples correctly classified
- Why does this make sense? Use same reasoning as logistic regression gradient.
- Say $y_i = 1$ and $\hat{y} = 0$. Then, after the update $\mathbf{w}^T \mathbf{x}_i$ gets bigger.

The Perceptron Learning Rule

- Amazing fact: If the data is linearly separable, the above algorithm always converges to a weight vector that separates the data.
- If the data is not separable, algorithm does not converge. Need to somehow pick which weight vector to go with.
- There are ways to do this (not examinable), such as the *averaged perceptron* and *voted perceptron*.
- This algorithm is a bit old and frumpy, but can still be very useful. Especially when you add kernels, to get the *kernel perceptron* algorithm. We may describe this later.
- Also can be seen as a very simple neural network, as we may also see later.