CSC321 Lecture 4
The Perceptron Algorithm

Roger Grosse and Nitish Srivastava

January 17, 2017
Recap: Perceptron Model

Inputs: \( x \).
Parameters: \( w \).

\[
y = \begin{cases} 
1 & \text{if } w^T x \geq 0 \\
0 & \text{otherwise}
\end{cases}
\]

An example of a binary linear classifier.
Recap: Perceptron Model

Inputs: $x$.
Parameters: $w$.

$$y = \begin{cases} 
1 & \text{if } w^T x \geq 0 \\
0 & \text{otherwise}
\end{cases}$$

An example of a binary linear classifier.

- Binary: Two possible classification decisions (0 or 1).
Recap: Perceptron Model

Inputs: \( x \).
Parameters: \( w \).

\[
y = \begin{cases} 
1 & \text{if } w^T x \geq 0 \\
0 & \text{otherwise}
\end{cases}
\]

An example of a binary linear classifier.

- Binary: Two possible classification decisions (0 or 1).
- Linear: \( w^T x \).
Recap: Perceptron Learning Algorithm

\[ w \leftarrow 0 \]

Repeat until all data points are classified correctly:

Choose a data point \( x \) with target \( t \)
Compute

\[ y = \begin{cases} 
1 & \text{if } w^T x \geq 0 \\
0 & \text{otherwise}
\end{cases} \]

If \( y \neq t \), then update

\[ w \leftarrow w + (t - y)x \]

Theoretical guarantee: if the data are linearly separable, it will make only a finite number of mistakes, then find a \( w \) which correctly classifies all training cases.

Note: after giving this lecture, we realized we’ve been inconsistent about what happens when an input lies on the decision boundary \( w^T x = 0 \). This isn’t a case we want to emphasize in this course. We won’t ask any exam or homework questions where inputs lie on the decision boundary. Sorry for the confusion.
Question 1: Perceptron example

Suppose we have the following data points, and no bias term:

- $x = (1, -2), t = 1$
- $x = (0, -1), t = 0$

The initial weight vector is $(0, -2)$.

- Draw the feasible regions in weight space.
  - Draw the axes in weight space $w_1, w_2$.
  - Draw each data point as a line that separates “good” and “bad” regions.
  - Shade the feasible region.

- Carry out the perceptron algorithm until you get a feasible solution.
  - It’s easiest to do it on the plot you made. Here is the algorithm -
    Choose a data point $x$ with target $t$
    Compute

$$y = \begin{cases} 
1 & \text{if } w^T x \geq 0 \\
0 & \text{otherwise}
\end{cases}$$

If $y \neq t$, then update

$$w \leftarrow w + (t - y)x$$
Question 2: Feature space

We’re given a problem with a single input and no bias parameter:
- $x = -1, t = 1$
- $x = 1, t = 0$
- $x = 3, t = 1$

Sketch the data in input space. Is this dataset linearly separable?
Question 2: Feature space

We’re given a problem with a single input and no bias parameter:
- $x = -1, \ t = 1$
- $x = 1, \ t = 0$
- $x = 3, \ t = 1$

Sketch the data in input space. Is this dataset linearly separable?

Design 2 basis functions (features) $\phi_1$ and $\phi_2$ such that
- $(\phi_1(-1), \phi_2(-1)), \ t = 1$
- $(\phi_1(1), \phi_2(1)), \ t = 0$
- $(\phi_1(3), \phi_2(3)), \ t = 1$

becomes linearly separable.

$$y = \begin{cases} 1 & \text{if } \mathbf{w}^T \Phi(x) \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Sketch the feature space - axes will be $\phi_1(x)$ and $\phi_2(x)$. 
Question 2: Feature space

We’re given a problem with a single input and no bias parameter:

- $x = -1, t = 1$
- $x = 1, t = 0$
- $x = 3, t = 1$

Sketch the data in input space. Is this dataset linearly separable?

Design 2 basis functions (features) $\phi_1$ and $\phi_2$ such that

- $(\phi_1(-1), \phi_2(-1)), t = 1$
- $(\phi_1(1), \phi_2(1)), t = 0$
- $(\phi_1(3), \phi_2(3)), t = 1$

becomes linearly separable.

$$y = \begin{cases} 
1 & \text{if } w^T \Phi(x) \geq 0 \\
0 & \text{otherwise}
\end{cases}$$

Sketch the feature space - axes will be $\phi_1(x)$ and $\phi_2(x)$.

Sketch the constraints in weight space.
Question 3: Linear regression in weight space

Recall that linear regression fits the model -

\[ y = wx + b. \]

Suppose we’re given the following training examples:

\[ x = -1, \ t = -1 \quad x = 0, \ t = 1 \quad x = 1, \ t = 2 \]

The optimal solution is (approximately) \( w = 1.5 \), \( b = 0.67 \).

For each example, sketch the sets of points in weight space which predict each target exactly, and plot the optimal solution. (The axes are \( b \) and \( w \).) What do you notice?
Question 3: Linear regression in weight space
What linear classifiers can’t represent.

Recall that Geoff said perceptrons can’t distinguish between two different binary patterns with wrap-around if they have the same number of nonzero entries.

Here’s another way of looking at it.

• Show that if a linear classifier classifies all the inputs $x^{(1)}, \ldots, x^{(N)}$ the same, then it also classifies their average the same.
• What is the average input for patterns A and B?
Your questions from the quiz

- How to initialize weights and biases?
  - by default, initialize to 0; but this depends on the situation
Your questions from the quiz

- How to initialize weights and biases?
  - by default, initialize to 0; but this depends on the situation

- Perceptrons with something other than a binary threshold?
  - We will cover neural net models which make other types of predictions (and these are sometimes called perceptrons)
Thinking about high-dimensional spaces

Geoff says that to think about 14-D space, you should “think about 3-D space and say 14 really loudly.” But some intuitions don’t carry over:

- “Most” sets of $D$ points in $D$ dimensions are linearly separable.
- “Most” points (inside a hypercube, say) are about the same distance from each other.
- “Most” vectors are approximately orthogonal to each other.