Reverse mode differentiation

Piece of computation:

Use standard rules:

\[ C = \cos A \implies \bar{A} = \bar{C} \odot \sin A \]
\[ C = AB \implies \bar{A} = \bar{C} B, \quad \bar{B} = A^T \bar{C} \]
\[ C = A + B \implies \bar{A} = \bar{C}, \quad \bar{B} = \bar{C} \]
\[ C = A^T \implies \bar{A} = \bar{C}^T \]

Multiple children

Apply rules separately for children and add
Matrix multiplication

\[ C = A \times B \quad O(LMN) \]

\[ C_{kn} = \sum_m A_{km} B_{mn} \]

LN terms \( \rightarrow \) cost \( O(M) \) each

Square matrix-matrix multiply \( O(N^3) \)

There are \( O(N^2) \) numbers in the matrices

[See also tutorial 5, Q1]
Autoencoder (Unsupervised)

Learning task
\[ f \approx x \]

not useful

[def autoencode(x):
    return x
    h = np.dot(I, x)
    f = np.dot(I, h)
    return f]

Dimensionality Reduction

D outputs
K units \( K \ll D \)

KxD inputs

\[ h = g^{(1)}(W^{(1)}x + b^{(1)}) \]
\[ f = g^{(2)}(W^{(2)}h + b^{(2)}) \] \leftarrow \text{decoder}

Encoder could be useful

Train this thing on a huge dataset
Use \( h(x; W^{(1)}, b^{(1)}) \) to transform data

2019 L22 3
Visualization

Set $k=2$

Contours of an RBF

$\phi_1$

$\phi_2$

$\phi_3$

We might want to increase dim. of data

$\phi_2$

$\phi_3$

$\phi_3$

weight vector for logistic regression.
Sparse Autoencoder

Encourage most elements of $h$ to be close to zero - sparse

Denoising Autoencoder

While training, we mask out (delete) some of the inputs, set $x^i$ to zero some

$m$ mask vector, of random 0's and 1's

Cost on an example

$$\| f(x^{(n)} \odot m) - x^{(n)} \|^2$$

Cost function

$$\sum_m p(m) \frac{1}{N} \sum_{n=1}^{N} \| f(x^{(n)} \odot m) - x^{(n)} \|^2$$

Monte Carlo

$\sim$ pick random $m$

$\sim$ random $n$
Principal Components Analysis (PCA)

Linear auto-encoder \( g''(a) = g''(a) = a \)

\[
h = V^T(x - \bar{x})
\]

\[
x = \left( Vh + \bar{x} \right).
\]

Training set mean

shared \( D \times k \) matrix

PCA advantages

- Fit columns of \( V \) to be eigenvectors of the covariance of data
  (No SGD!)

- Same answer every time.

- The solutions for different \( k \)
  they're nested

\[
h_1(x) \text{ it's the same for all } k
\]

\[
h_2(x) \text{ " " " " " } k \geq 2
\]
PCA applied to bodies

$\mathbf{e}_1 \quad \mathbf{e}_2 \quad \mathbf{e}_3 \quad \mathbf{e}_4 \quad \mathbf{e}_5$

$\mu \quad +4\sigma \quad -4\sigma$

Freifeld and Black, ECCV 2012
PCA applied to DNA

Novembre et al. (2008) — doi:10.1038/nature07331
Carefully selected both individuals and features

1,387 individuals

197,146 single nucleotide polymorphisms (SNPs)

Each person reduced to two(!) numbers with PCA
MSc course enrollment data

Binary $S \times C$ matrix $M$

$M_{sc} = 1$, if student $s$ taking course $c$

Each course is a length $S$ vector

... OR each student is a length $C$ vector
PCA applied to MSc students