50.039 – Deep Learning

Alex

Week 03: Overfitting and convolutions

[The following notes are compiled from various sources such as textbooks, lecture materials, Web resources and are shared for academic purposes only, intended for use by students registered for a specific course. In the interest of brevity, every source is not cited. The compiler of these notes gratefully acknowledges all such sources.]

1 Some Einsum

Which einsum notation is required to implement the following operations? Remember it is a pair

$$indices_1, indices_2, indices_3, \ldots, indices_n - > indices_r, [t_1, t_2, t_3, \ldots, t_n]$$

- $$C_{j,k} = \sum_i A_{ijk} b_i$$
- $$C_j = \sum_{i,k} A_{ijk} b_{ik}$$
- $$A_{ik} = \sum_{j,l} A_{ijkl}$$
- yes this is not the same as before, note the change in index ordering
$$A_{ki} = \sum_{j,l} A_{ijkl}$$
- $$C_i = \sum_{j,k} A_{ijk} A_{ijk}$$

- $$C = x^\top A x, x \in \mathbb{R}^d,\ 1-tensor, A \in \mathbb{R}^{d\times d},\ 2-tensor,$$

- $$C = AG^\top B, A \in \mathbb{R}^{d\times e},\ 2-tensor, G \in \mathbb{R}^{f\times e},\ 2-tensor, B \in \mathbb{R}^{f\times l},\ 2-tensor,$$

  The result is a tensor of what order here ? in any case there is more than one possible output index ordering in the sense of C_{ijk} vs C_{jki} vs C_{kij} and so on . any output index ordering is okay here

- $$C_{????} = \sum_{cd} A_{abcd} B_{bcde} E_{cdef}$$

  any output index ordering is okay here again

2 Overfitting with more and more dimensions

Lets consider the case when we have a fixed number of datapoints n and we go into more and more high dimensional spaces.
More precisely:

- we have a classification problem with samples (x, y) with $y \in \{-1, +1\}$ being the labels.

- Suppose for now that we have a one-dimensional feature $x_i = (x_i^{(1)})$ where $x^{(1)}$ denotes the index for the only dimension, and the subscript $_i$ in x_i is the number of the sample. I introduce this notation, because we will consider soon samples in D dimensions $x_i = (x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(D)})$. Consider the following distribution of samples.

  $$P(X^{(1)} < 0|Y = -1) = 0.5$$
  $$P(X^{(1)} < 0|Y = +1) = 0.5$$

  This tells that the classifier

  $$f_0(x) = 2I[x^{(1)} \geq 0] - 1 = \begin{cases} -1 & x^{(1)} < 0 \\ +1 & x^{(1)} \geq 0 \end{cases}$$

  is not that excessively useful as a predictor under the expectation under $P(x, y)$.

- compute $E_{(x,y)\sim P}[I[f_0(x) \neq y]]$. Show your work in detail. This works for any value of $P(Y = +1)$.

- Suppose we draw the N samples statistically independently. Let the first $N/2$ points be of class -1.

  What is the probability that we draw N samples such that the error on this training dataset is zero under $f_0(x)$? Express this event in terms of conditions to x_i for the first $N/2$ points and for the last $N/2$ points. Then compute its probability under above $P(X|Y)$.

- now lets consider a D-dimensional setup. $x_i = (x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(D)})$

$$P(X^{(d)} < 0|Y = -1) = 0.5 \ \forall d = 1, \ldots, D$$
$$P(X^{(d)} < 0|Y = +1) = 0.5 \ \forall d = 1, \ldots, D$$

  and all the dimensions are statistically independent, thus e.g.

$$P(X^{(d_1)} < 0, X^{(d_2)} < 0, X^{(d_3)} < 0|Y) = \prod_{k=1}^{3} P(X^{(d_k)} < 0|Y)$$

  From the $D = 1$ case above you know the distribution of the case when in one of these D dimensions the error on this training dataset is zero under $f_0(x^d)$.

- What is the probability distribution that we draw N samples such that in exactly K out of D dimensions (remember $x_i = (x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(D)})$) $\{d_1, \ldots, d_K\} \subset 1, \ldots, D$ $f_0(x^{(d_k)})$ achieves zero training error? Give its name and its parameters.

- What is the precise probability that we draw N samples such that in at least one dimension d out of D dimensions $f_0(x^{(d)})$ achieves zero training error?

- What is the limit of this probability as $D \to \infty$? What is the $\mathcal{O}(\cdot)$ complexity of the convergence of this limit as a function of D ?

  Hope that tells you something about spurious correlations in high dimensions.