Assignment01

May 7, 2018

Write by Zhao, Xiaoning and Zhang, Guangde

1 Q 1 Advances in neural network technology

The 3 most important factors that advance modern deep neural networks are:

- Increasing dataset sizes
- Increasing accuracy
- Increasing real-world impact

The reason are:

1. Increasing dataset sizes. The important thing with deep learning and machine learning in general is it needs a lot of data to train on. The dataset size is tightly coupled with the heart of the induction problem that we are trying to solve in modern DNN. It is also related to the complexity of the problem and the learning algorithm. Thus, I think it's quite an important factor.

2. Increasing accuracy. With a small test set, 100% accuracy is possible, however, According to Bayes Error, as we continue to add more test samples, the accuracy will be reduced. I think it's worth to work on it and improve that.

3. Increasing real-world impact. It took about 5 years to get neural nets to show impressive results, first in speech recognition, then computer vision, and more recently in natural language processing, the resulting algorithms have sparked a revolution in academic and industrial applications. Thus, the real-world impact will be a hot topic I assume.

2 Q 2 Eigenvectors and eigenvalues

2.1 a

$\because A \in R^{nxn}$ is diagonalizable, $A = T\Lambda T^{-1}$ for an invertible matrix $T \in R^{nxn}$

$Ax = \lambda x$

$AT = T\Lambda T^{-1}T = T\Lambda I = T\Lambda$

$\therefore$ the eigenvector/eigenvalue pairs of A are $(t^{(i)}, \lambda_i)$

2.2 b

$\because B = UU^{-1}$ and Eigendecompostion of B is $B = Udiag(\lambda)U^{-1}$
$\therefore B^T = (UVU^{-1})^T$
$= U^{TT}(U^{-1})^T$
$= U^T(U^T)^T$
$= U^TU$
$= U^{-1}U$
$= B$

2.3 c

$\because$ As given matrix A is S_+^n
$\therefore \forall z,\, z^TA \geq 0$
Let λ be a eigenvalue of A and x be a corresponding real *eigenvector*, then we have

$$Ax = \lambda X$$

The we multiply by x^T on left and obtain:

$$x^TAx = \lambda x^Tx = \lambda||x||^2$$

The left part of the equation is positive (≥ 0)
x is nonzero vector
since the length $||x||^2 \geq 0$
$\therefore \lambda$ must be non-negative.

3 Q 3 Trace operator, matrix derivatives

3.1 a) $\nabla tr(AB) = B^T$

$\because tr(AB) = \sum_k A_{ik}B_{kj} = \sum_{k=1}^k A_{1k}B_{k1} + \sum_{k=1}^k A_{2k}B_{k2} + ... + \sum_{k=1}^k A_{ik}B_{kj}$
$\therefore \frac{\nabla tr(AB)}{\partial A_{ij}} = B_{ji}$
$\therefore \nabla tr(AB) = B^T$

3.2 b) $\nabla_{A^T} f(A) = (\nabla_A f(A))^T$

$$\nabla_{A^T} f(A) = \begin{bmatrix} \frac{\partial f(A)}{\partial A_{11}} \cdots \frac{\partial f(A)}{\partial A_{n1}} \\ \frac{\partial f(A)}{\partial A_{12}} \ddots \frac{\partial f(A)}{A_{n2}} \\ \vdots \\ \frac{\partial f(A)}{\partial A_{1n}} \cdots \frac{\partial f(A)}{\partial A_{nn}} \end{bmatrix}$$
$= (\nabla_{A^T} f(A))^T$

3.3 c) $\nabla_A tr(ABA^TC) = CAB + C^TAB^T$

Let's have $AB = f(A)$, where f is matrix-valued

$\nabla_A tr(ABA^TC) = \nabla_A tr(f(A)A^TC)$
$= \nabla_x tr(f(x)A^TC) + \nabla_x tr(f(A)x^TC)$
$= (A^TC)^T f'(x) + (\nabla_{x^T} tr(f(A)x^TC))^T$
$= C^TAB^T + (\nabla_{x^T} tr(x^TCf(A)))^T$
$= C^TAB^T + ((cf(A))^T)^T$
$= C^TAB^T + CAB$

3.4 d) If $z = Ax$, then $\nabla_A f(z) = (\nabla_z f(z)) \otimes x$

$\because z = Ax$
$\nabla_A f(z) = \frac{\partial f(z) \partial z}{\partial A}$
$\nabla_z f(z) = 2z$
$(\nabla_z f(z)) \otimes x = (\nabla_z f(z))x^T = 2zx^T$