Homework (2)

1. For each of the following sequences, determine the rate of convergence and the rate constant.

   a. $x_k = 2^{-k}$, for $k = 1, 2, \cdots$.

   b. $x_k = 1 + 5 \times 10^{-2k}$, for $k = 1, 2, \cdots$.

   c. $x_k = 2^{-2^k}$.

   d. $x_k = 3^{-k^2}$.

   e. $x_k = 1 - 2^{-2^k}$ for k odd, and $x = 1 + 2^{-k}$ for k even.

2. Let $\{x_k\}$ and $\{c_k\}$ be convergent sequences, and assume that

$$\lim_{k \to \infty} c_k = c \neq 0.$$

Consider the sequence $\{y_k\}$ with $y_k = c_k x_k$. Can its convergence rate and rate constant be determined from those of $\{x_k\}$ and $\{c_k\}$?

Homework (2)

3. Compute the gradient and Hessian of the following functions:

   a. $f(\mathbf{x}) = \|\mathbf{x}\|_p$, $\mathbf{x} \neq \mathbf{0}$, $p > 1$.

   b. $f(\mathbf{x}) = (\mathbf{a}^T\mathbf{x})(\mathbf{b}^T\mathbf{x})$.

   c. $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$.

Homework (2)

4. Compute the gradient of the following matrix functions (write in matrices, rather than entrywise. Give details. Working out 5 is sufficient if you don't want to torture yourself.):

a. $f(\mathbf{X}) = \|\mathbf{X}\|_F$.

b. $f(\mathbf{X}) = \|\mathbf{X}^T\mathbf{A}\mathbf{X}\|_F^2$ and $\mathbf{A}$ is a symmetric matrix.

c. $f(\mathbf{X}) = \|\operatorname{diag}(\mathbf{X}^T\mathbf{A}\mathbf{X})\|_F^2$ and $\mathbf{A}$ is a symmetric matrix.

d. $f(\mathbf{X}) = \det \mathbf{X}$.

e. $f(\mathbf{X}) = \det(\mathbf{X}^T\mathbf{A}\mathbf{X})$.

f. $f(\mathbf{X}) = \mathbf{a}^T\mathbf{X}^{-1}\mathbf{b}$.

g. $f(\mathbf{X}) = \operatorname{tr}(\mathbf{A}\mathbf{X}\mathbf{B})$.

h. $f(\mathbf{X}) = \operatorname{tr}(\mathbf{A}\mathbf{X}^{-1}\mathbf{B})$.

i. $f(\mathbf{X}) = \operatorname{tr}((\mathbf{A}+\mathbf{X})^{-1})$.

Homework (2)

5. Write a program for computing the gradient of error function V with respect to the weights of a neural network with arbitrary topology. Test it on simplified ResNet: 1 filter, 1 skip connection, no pooling and batch normalization. Verify the correctness by numerical differentiation, i.e. verify

$$\frac{V(\mathbf{w}+t\mathbf{v})-V(\mathbf{w})}{t} \approx \langle \nabla V(\mathbf{w}), \mathbf{v} \rangle$$

for arbitrary choice of $\mathbf{v}$ and sufficiently small t. The code should include the verification step.

6. Write an automatic differentiation program that works on a given expression DAG. Test it on function y_o and verify it with numerical differentiation:

$$y_o = (\sin(x_1+1)+\cos(2x_2))\tan(\log(x_3))+(\sin(x_2+1)+\cos(2x_1))\exp(1+\sin(x_3)).$$

The code should include the verification step.

Homework (2)

7. Find the dual norm of Mahalanobis norm: $\|\mathbf{x}\|_{\mathbf{M}} = \sqrt{\mathbf{x}^T\mathbf{M}\mathbf{x}}$, where $\mathbf{M}$ is a positive definite matrix.
8. Prove that the eigenvalues λ_i of $(\mathbf{A}+\mathbf{B})^{-1}\mathbf{A}$, where $\mathbf{A}$ is positive semidefinite and $\mathbf{B}$ is positive definite, satisfy $0 \leq \lambda_i < 1$.
9. Compute the condition number of the following matrix:

$$\begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 5 & 6 & 9 \end{bmatrix}.$$

10. Suppose $\mathbf{X} \in \mathbb{R}^{3\times 3}$, $\mathcal{A}(\mathbf{X}) = X_{11} + X_{12} - X_{31} + 2X_{33}$, find $\mathcal{A}^*$.