The Exercise of Week 1

Name: Han Siyue

ID: 17307110012

September 14, 2019

1 Basic Operation

using the definitions, $\alpha = 8,\ x = \begin{bmatrix} 1 \\ 3 \end{bmatrix},\ y = \begin{bmatrix} 2 \\ 4 \end{bmatrix},\ A = \begin{bmatrix} 1 & 3 \\ 3 & 2 \end{bmatrix}$ to evaluate the following expressions:

1. $\alpha(x+y) = 8\begin{bmatrix} 1+2 \\ 3+4 \end{bmatrix} = \begin{bmatrix} 24 \\ 56 \end{bmatrix}$

2. $x^T y + \parallel x \parallel^2 = 1 \times 2 + 3 \times 4 + 1 \times 1 + 3 \times 3 = 24$

3. $Ax = \begin{bmatrix} 1 \times 1 + 3 \times 3 \\ 3 \times 1 + 2 \times 3 \end{bmatrix} = \begin{bmatrix} 10 \\ 9 \end{bmatrix}$

4. $A^T A = \begin{bmatrix} 1 \times 1 + 3 \times 3 & 1 \times 3 + 3 \times 2 \\ 3 \times 1 + 2 \times 3 & 3 \times 3 + 2 \times 2 \end{bmatrix} = \begin{bmatrix} 10 & 9 \\ 9 & 13 \end{bmatrix}$

5. $Tr(A^T A) = 10 + 13 = 23$

If {x, y, z} are real-valued column-vectors of length d and {A, B, C} are real-valued d by d matrices, state whether each of the below statements is true or false in general; if the statement is false,please give the correct form.

1. $x(y+z)^T = z^T + y^T x$ is false, and the correct form should be $x(y+z)^T = x(z^T + y^T)$

2. $x^T x = x x^T$ is false, and does not have a similar form which is correct.

3. $x^T A y = y^T A^T x$ is false, and the correct form should be $(x^T A y)^T = y^T A^T x$

4. $x^T y^T z = (xy)^T z$ is false, and the correct form should be $x^T y^T z = (yx)^T z$

5. $AB = BA$ is false, and does not have a similar form which is correct.

6. $(AB)C = A(BC)$ is true.

7. $A(B+C)^T = B^T A + C^T A$ is false, and the correct form should be $A(B+C)^T = AB^T + AC^T$

8. $(AB)^T = A^T B^T$ is false, and the correct form should be $(AB)^T = B^T A^T$

2 Gradients and Hessians of Linear and Quadratic Functions

we use the convention that all values are real and :

1. α is a scalar.
2. a and b are length-d column-vectors.
3. Element i of b is denoted by b_i.
4. A and B are d by d matrices.
5. Row i of A is denoted by a_i^T.
6. W is a symmetric d by d matrix.

Express the gradient $\nabla f(x)$ and Hessian $\nabla^2 f(x)$ of the following linear/quadratic functions in matrix notation, simplifying as much as possible.

1. $f(x) = a^T x + \alpha$ (linear)

   $\nabla f(x) = a$

   $\nabla^2 f(x) =$**0** , in which **0** represents a d by d matrix that all the elements are 0.

2. $f(x) = a^T x + a^T Ax + x^T A^T b$ (more linear forms)

   $\nabla f(x) = a + A^T a + A^T b$

   $\nabla^2 f(x) =$**0** , in which **0** represents a d by d matrix that all the elements are 0.

3. $f(x) = x^T x + x^T Wx + x^T ABx$ (quadratic forms)

   $\nabla f(x) = 2x + 2Wx + 2ABx$

   $\nabla^2 f(x) = 2W + 2AB + 2I_d$

4. $f(x) = \frac{1}{2}(Ax-b)^T W(Ax-b)$ (weighted least squares)

   $\nabla f(x) = A^T W(Ax-b)$

   $\nabla^2 f(x) = A^T WA$

5. $f(x) = \frac{\lambda}{2}\| x \|^2 + \frac{1}{2}\sum_{i=1}^{n}(a_i^T x - b_i)^2$ (L2-regularized least squares)

   $\nabla f(x) = \lambda x + \sum_{i=1}^{n} a_i(a_i^T x - b_i) = \lambda x + A^T(Ax-b)$

   $\nabla^2 f(x) = \lambda I_d + A^T A$