James Zhao
HW3
A15939512

1 1

$\frac{x}{\|x\|} = \frac{(1,2,3)}{\sqrt{1+4+9}} = (\frac{1}{\sqrt{14}}, \frac{2}{\sqrt{14}}, \frac{3}{\sqrt{14}})$

2 2

If a vector v is orthogonal to $(1,1)$, then $(1,1) \cdot v = 0$, $v_1 + v_2 = 0, v_1 = -v_2$. Thus, the vector is parallel to $(1,-1)$. The only unit vectors satisfying this are the unit vector and negative unit vector, or $(\sqrt{2}/2, -\sqrt{2}/2)$ and $(-\sqrt{2}/2, \sqrt{2}/2)$.

3 3

$(x_1, x_2) \cdot (x_1, x_2) = 25$
$x_1^2 + x_2^2 = 25$
I would describe it as a circle centered at the origin with radius 5.

4 4

$f(x) = (w_1, w_2, w_3) \cdot (x_1, x_2, x_3)$
$f(x) = (2, -1, 6) \cdot (x_1, x_2, x_3)$
$w = (2, -1, 6)$

5 5

When doing matrix multiplication, $(m,k) \times (k,n) \rightarrow (m,n)$. If A has 30 columns, then $k = 30$. The output $(m,n) = (10,20)$, so $m = 10$,$n = 20$. Thus, A is a 10x30 matrix, B is a 30x20 matrix.

6 6

6.1 a

$n \times d$

6.2 b

$(n,d) \times (d,n) \rightarrow (n,n)$
Answer:
$n \times n$

6.3 c

It is the i'th data point dot-producted with the j'th data point

7 7

Assuming a convention that x is a column-vector:
$x^T x = \|x\|^2$
$x^T x x^T x x^T x = \|x\|^6$

8 8

8.1 $x^T x$

$(1, 3, 5) \cdot (1, 3, 5) = 1 * 1 + 3 * 3 + 5 * 5 = 1 + 9 + 25 = 35$

8.2 xx^T

$\begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} \begin{bmatrix} 1 & 3 & 5 \end{bmatrix}$
$\begin{bmatrix} 1 & 3 & 5 \\ 3 & 9 & 15 \\ 5 & 15 & 25 \end{bmatrix}$

9 9

$\|x\| = 2, \|y\| = 2, x \cdot y = 2$
$\cos\theta = \frac{x \cdot y}{\|x\|\|y\|}$
$\cos\theta = \frac{2}{2*2} = \frac{1}{2}$
$\theta = \cos^{-1}(0.5) = \frac{\pi}{3}$

10 10

$$x^T \begin{bmatrix} 3 & 1 & -2 \\ 1 & 0 & 0 \\ -2 & 0 & 6 \end{bmatrix} x$$

11 11

11.1 a

AA^T must be symmetric since the i, j entry $(A_{i,j})$ are the dot products of the i'th row and j'th row. Dot product is symmetric, so the $A_{i,j} = A_{j,i} = A_{i,:} \cdot A_{j,:}$

11.2 b

$A^T A$ must be symmetric since it is is the same as $A' = A^T, (A' A'^T)$, and we just showed that AA^T is necessarily symmetric.

11.3 c

$A + A^T$ is necessarily symmetric. For some pair $(i, j), i \neq j$ of the resulting matrix M, we can check $M_{i,j} = A_{i,j} + A^T_{i,j} = A_{i,j} + A_{j,i}$, $M_{j,i} = A_{j,i} + A^T_{j,i} = A_{j,i} + A_{i,j} = M_{i,j}$. Because transposed-corresponding off-diagonal elements are equivalent, the matrix $A + A^T$ must be necessarily symmetric.

11.4 d

$A - A^T$ is not necessarily symmetric. Let us craft a counterexample:
$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$
$A - A^T = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$
which is clearly not symmetric.

12 12

12.1 a

Determinant of a diagonal matrix is the product of entries, so it is $8! = 40320$

12.2 b

Inverse of a diagonal matrix is a matrix of the diagonal elemenets inverted (according to lecture), so it is $diag(\frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \frac{1}{7}, \frac{1}{8})$

13 13

13.1 a

If the matrix U is square, has rows of orthonormal vectors, then it is an orthogonal matrix. UU^T is a matrix where the i, j'th entry is a dot product of the i'th and j'th vector. The dot product of a unit vector with itself is 1, so the diagonal entries are all 1. Because the rows form an orthonormal basis, all $U_{i,:} \cdot U_{j,:}, i \neq j = 0$, thus the non-diagonal entries are all zeros. This results in $UU^T = I$.

13.2 b

We can trivially show that $UU^T = I = I^T = (UU^T)^T = U^T U^{TT} = U^T U$. For any invertible square matrix, we also know that $AA^{-1} = A^{-1}A = I$. Matching these equations together, we can see that $U^T = U^{-1}$.

14 14

$\det(A) = 1 * z - 2 * 3 = 0$
$z = 6$

15 15

15.1 a

Training Procedure/Overview:
The dataset was split by randomly sampling 50000 indexes out of the integers in the range [0, 59999] using np.random.choice without replacement, and those indexes correspond to the (sample, label) tuples that comprise the training set. The remaining 10000 non-sampled indexes comprise the validation set.
Next, the code supplied by the jupyter notebook was used to test ranges of values of c, evaluating on the validation set (rather than the test set, which the code was originally set up to do)

- First, a wide variety of power-of-2 c-values were tested, ranging from small powers of 2 to large powers of 2 (min: 2^{-2}, max: 2^{17})
- After finding the best values from the initial broad search (I determined it was from [1024,8192]), a fine=grained search over that interval is performed, sampling 20 equally-spaced points in the interval. From here, the c that resulted in the lowest validation error was chosen as the best hyper-parameter value and was used to evaluate the test set.

Lastly, after finding the best value of c, the model is run through the test set using the method provided by the starter code.
Code will be applied to this question's submission.

15.2 b

c=3664

15.3 c

0.0438 Error Rate on test set

15.4 d

See attached submission pages