Homework 1 Solutions

1.1 (c,d)

(a) Affine. We prove this by demonstrating that it can be written in the affine form $f(x) = a^T x + b$ (slide **1-30**).

Expanding the norms,

$$\begin{aligned} f(x) &= \|x - c\|^2 - \|x - d\|^2 \\ &= (x - c)^T (x - c) - (x - d)^T (x - d) \\ &= (x^T x - 2c^T x - x^T c + c^T c) - (x^T x - d^T x - x^T d - d^T d) \\ &= 2(d - c)^T x + \|c\|^2 - \|d\|^2 \end{aligned}$$

Therefore f can be written in the form $f(x) = a^T x + b$ where

$$a = 2(d - c), \qquad b = \|c\|^2 - \|d\|^2$$

(b) Not linear or affine. We prove this by showing that f violates the definition of an affine function (slide **1-29**)

As an example,

$$x = \begin{bmatrix} 0 \\ 1 \\ 4 \end{bmatrix}, \qquad x = \begin{bmatrix} 6 \\ 1 \\ 0 \end{bmatrix}, \qquad \alpha = \beta = 1/2$$

Substituting in these values,

$$\begin{aligned} \alpha f(x) + \beta f(y) &= 1 \\ f(\alpha x + \beta y) &= 2 \end{aligned}$$

Hence $\alpha f(x) + \beta f(y) \neq f(\alpha x + \beta y)$ and so the function is not affine. Because all linear functions are affine, the function is also not linear.

1.6

(a)

$$\begin{aligned} (a + b)^T (a - b) &= a^T (a - b) + b^T (a - b) \\ &= a^T a - a^T b + b^T a - b^T b \\ &= \|a\|^2 - \|b\|^2 \end{aligned}$$

(b)

$$\begin{aligned}\|a+b\|^2 + \|a-b\|^2 &= (a+b)^T(a+b) + (a-b)^T(a-b) \\ &= a^Ta + b^Ta + a^Tb + b^Tb + a^Ta - b^Ta - a^Tb + b^Tb \\ &= 2\|a\|^2 + 2\|b\|^2\end{aligned}$$

1.11

Define $y_i = 1/n$, $i = 1, \ldots, n$. Then

$$x^Ty = \sum_{i=1}^{n} x_i y_i = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \|y\| = \sqrt{\sum_{i=1}^{n} y_i^2} = \frac{1}{\sqrt{n}}$$

Substituting this into the Cauchy–Schwartz inequality

$$-\|x\|\|y\| \leq \sum_{i=1}^{n} x_i y_i \leq \|x\|\|y\|$$

we get

$$-\frac{1}{\sqrt{n}}\|x\| \leq \frac{1}{n}\sum_{i=1}^{n} x_i \leq \frac{1}{\sqrt{n}}\|x\|$$

The upper bound holds if and only if x and y are parallel. By our definition of y, this is true when every element of x is equal and nonnegative.

The lower bound holds if and only if x and y are antiparallel. Similarly, this is true when every x_i is equal and nonpositive.

1.22

```
load mnist_train;
digits = digits(:,1:10000);
[n, N] = size(digits);
K = 20;

class = randi(K, 1, N);
Z = zeros(n,K);
D = zeros(K, N);
Jprev= NaN;
for iter = 1:100
  for k = 1:K
    I = find(class == k);
    Z(:,k) = mean(digits(:,I), 2);
```

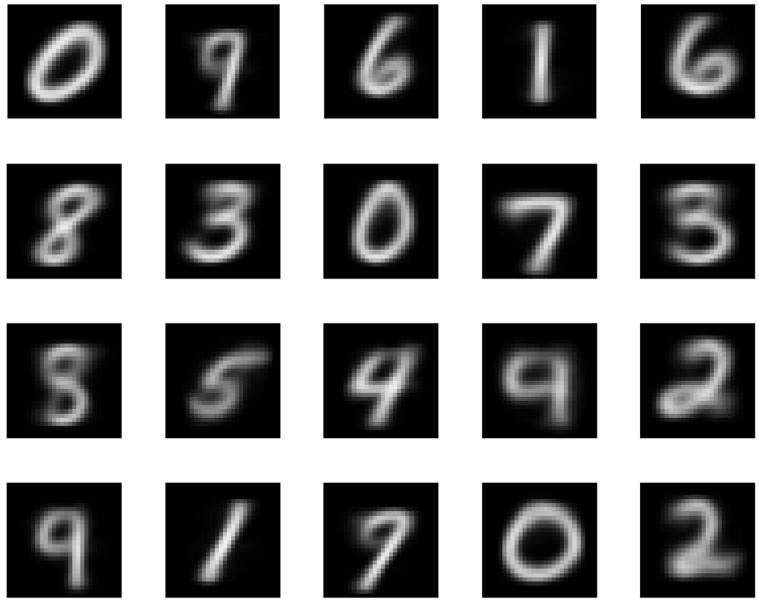

Figure 1: Typical result.

```
  end
  for k = 1:K
    D(k,:) = sqrt( sum( (digits - Z(:, k*ones(1,N))).$^$2) );
  end;
  [d, class] = min(D);
  J = (1/N) * norm(d)$^$2;
  if iter > 1
    if abs(J - Jprev) < 1e-5 * J, break; end;
    Jprev = J;
  end;
end;

for k=1:K
  subplot(4,5,k)
  imshow(reshape(Z(:,k), 28, 28));
end
```