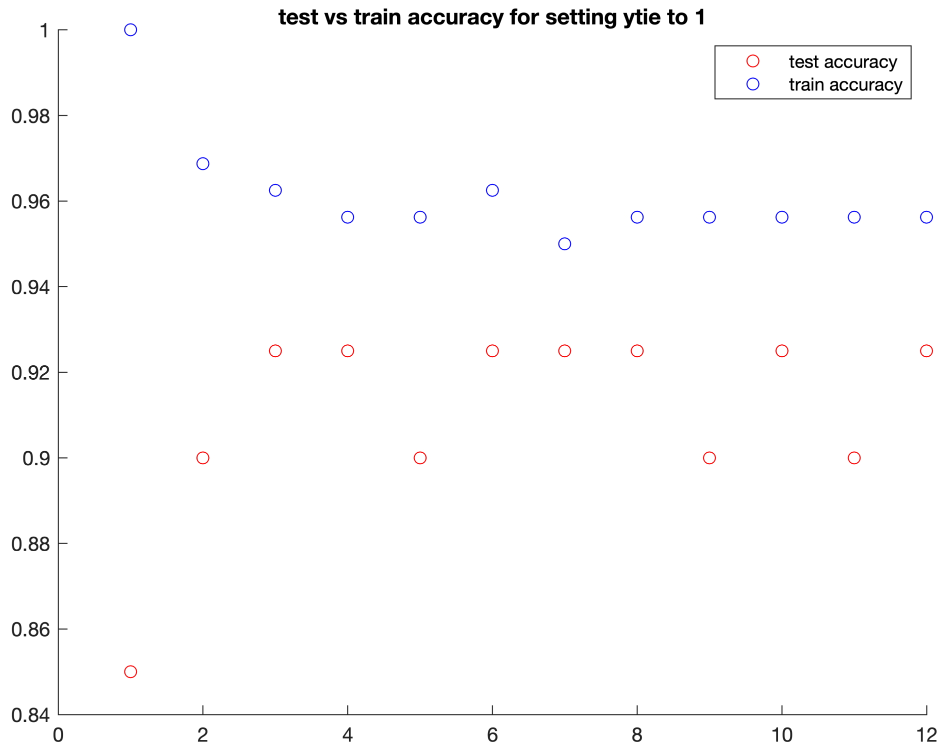
PROBLEM 1 ON HW 4

```
close all;
stat30 = readtable("UCLA_EE_grad_2030.csv");
stat31 = readtable("UCLA_EE_grad_2031.csv");
data = [stat30; stat31];
tes = table2array(data(1:40,:));
tra = table2array(data(41:end,:));
```

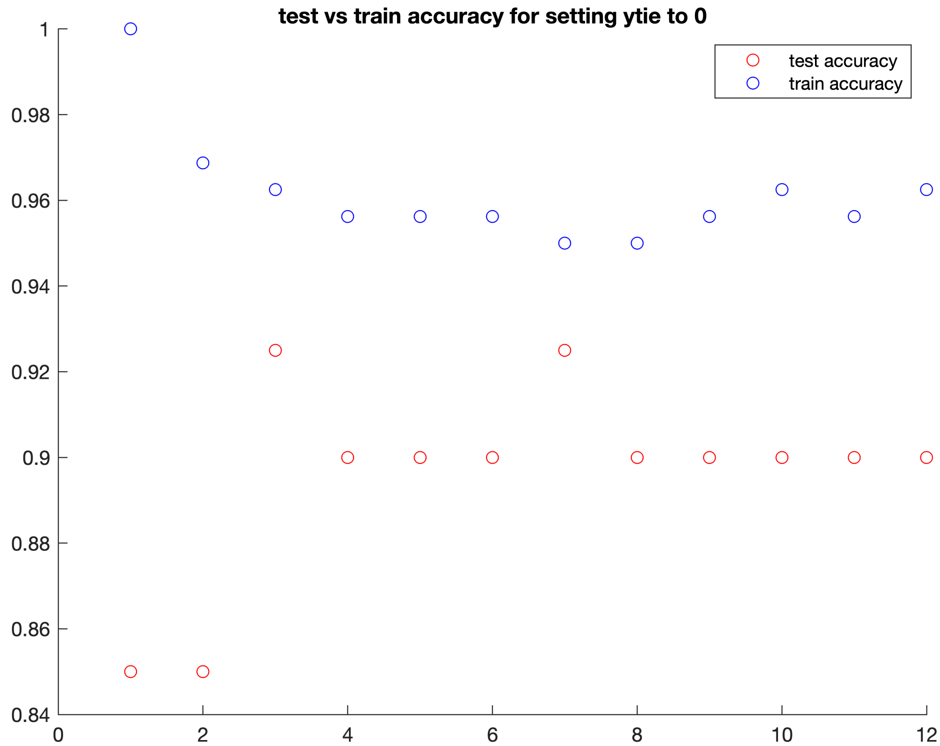
(a) $y_{tie} = 1$: implement k-NN algorithm and plot the training/testing accuracy for k = 1, 2, ...12.

```
testacc = zeros(2, 12);
traiacc = zeros(2, 12);
for k = 1:12
    testyay = [0;0];
    traiyay = [0;0];
    for i = 1:40
        d = sortrows([(tra(:,1)-tes(i,1)).^2+(tra(:,2)-tes(i,2)).^2 tra(:,3)]);
        y1 = sum(d(1:k,2));
        y2 = y1;
        if y1==0, y1 = 1; end
        if y2==0, y2 = -1; end
        y1 = y1 / abs(y1);
        y2 = y2 / abs(y2);
        if y1 == tes(i, 3)
            testyay(1) = testyay(1) + 1;
        end
        if y2 == tes(i, 3)
            testyay(2) = testyay(2) + 1;
        end
    end
    for i = 1:160
        d = sortrows([(tra(:,1)-tra(i,1)).^2+(tra(:,2)-tra(i,2)).^2 tra(:,3)]);
        y1 = sum(d(1:k,2));
        y2 = y1;
        if y1==0, y1 = 1; end
        if y2==0, y2 = -1; end
        y1 = y1 / abs(y1);
        y2 = y2 / abs(y2);
        if y1 == tra(i, 3)
            traiyay(1) = traiyay(1) + 1;
        end
        if y2 == tra(i, 3)
            traiyay(2) = traiyay(2) + 1;
        end
    end
    testacc(1,k) = testyay(1) / 40;
    testacc(2,k) = testyay(2) / 40;
    traiacc(1,k) = traiyay(1) / 160;
    traiacc(2,k) = traiyay(2) / 160;
end
figure(1);
hold on;
```

```
scatter((1:1:12), testacc(1,:)', 'red');
scatter((1:1:12), traiacc(1,:)', 'blue');
legend('test accuracy', 'train accuracy');
title('test vs train accuracy for setting ytie to 1');
```

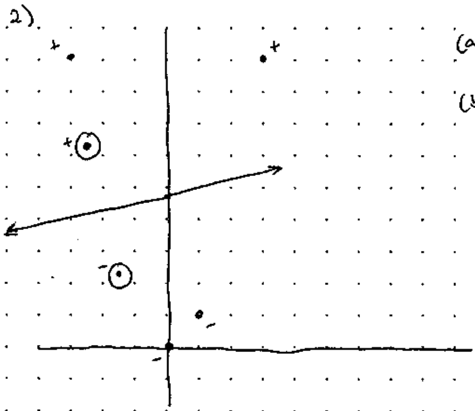



```
figure(2);
hold on;
scatter((1:1:12), testacc(2,:)', 'red');
scatter((1:1:12), traiacc(2,:)', 'blue');
legend('test accuracy', 'train accuracy');
title('test vs train accuracy for setting ytie to 0')
```



(c) The accuracy against training and testing data is actually fairly good! As k increases the training accuracy drops quite a bit in the beginning but levels out at around 0.96 or 96%. As k increases the testing data accuracy increases by quite a lot in the beginning but levels out at around 0.9 or 90%. A larger k shouldn't make a big difference after 3 or 4. Looking at the graph for ytie = 1 an interesting pattern results. When k is even, the accuracy is noticeably higher than when k is odd. For the ytie = 0 though it doesn't really make a difference. Odd and even make a big difference, though, because when the two classes appear equally. Deciding comes down to your choice. They aren't contradictory to each other. It seems that when ytie is set to 0, even values of k do better, meaning that it is more likely that it is a 1. Setting ytie to 0 should not make it do better when it is even, and so it's the same odd vs even.

2)



(a) $(-2.5, 6.25)$, $(-1.5, 2.25)$

(b) $(-2, 4.25)$

$m = -4 \qquad \perp m = \frac{1}{4}$

$y = \frac{1}{4}x + b$

$4.25 = -0.5 + b$

$b = 4.75$

$y = 0.25x + 4.75$

(c) $a_2 \cdot y^{(2)} \cdot x^T \cdot x^{(2)} + b$
$+ a_4 \cdot y^{(4)} \cdot x^T \cdot x^{(4)}$

$a_2 y^{(2)} + a_4 y^{(4)} = 0 \Rightarrow a_2 = a_4$

$y = w^T x + b$

$0 = \frac{2}{17}\begin{bmatrix} -1 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - 2.235 = 0$

$2[-x_1 + 4x_2] - 38 = 0$

$-x_1 + 4x_2 = 19$

$x_2 = 0.25x_1 + 4.75$ ✓

same equation.

$$L = (a_4 + a_2) - \frac{1}{2}\left[a_2^2 \|x^{(2)}\|^2 + a_4^2 \|x^{(4)}\|^2 - 2a_2 a_4 x^{(2)T} x^{(4)}\right]$$

$\|x^{(2)}\|^2 = 45.312 \qquad \|x^{(4)}\|^2 = 7.312 \qquad x^{(2)T} x^{(4)} = 17.813$

$2a_2 - \frac{1}{2} 17 \cdot a_2^2$

$$\frac{\partial L}{\partial a_2} = 2 - 17a_2 = 0 \rightarrow \boxed{a_2 = a_4 = \frac{2}{17}}$$

$$w = \sum_{n \in S} a_n y_n x_n = \frac{2}{17} \cdot 1 \cdot \begin{bmatrix} x_2 \end{bmatrix} + \frac{2}{17} \cdot (-1) \cdot \begin{bmatrix} x_4 \end{bmatrix}$$

$$= \boxed{\frac{1}{17}\begin{bmatrix} -2 \\ 8 \end{bmatrix}}$$

$$b = \frac{1}{|S|}\sum_{n \in S}\left[t_n - w^T x_n\right] = \frac{1}{2}\left[1 - \begin{bmatrix} -2/17 & 8/17 \end{bmatrix}\begin{bmatrix} -2.5 \\ 6.25 \end{bmatrix} + (-1) - \begin{bmatrix} -2/17 & 8/17 \end{bmatrix}\begin{bmatrix} -1.5 \\ 2.25 \end{bmatrix}\right]$$

$$= \boxed{-2.235}$$

PROBLEM 3 ON HW 4

```
close all;
stat = readtable("UCLA_EE_grad_2031.csv");
```

(a)

```
y = stat{:,3};
GPA = stat{:,1};
GRE = stat{:,2};
hold on;
scatter(GPA(y == -1), GRE(y == -1), 'blue');
scatter(GPA(y == 1), GRE(y == 1), 'red');
```

(b)

```
x = stat{:,1:2};
y = stat{:,3};
cvx_begin
    variable w(2)
    variable b
    minimize( 1/2 * sum_square_abs(w) )
    subject to
        for i = 1:100
            y(i) * (w' * x(i,:)' + b) >= 1
        end
cvx_end
```

```
Calling SDPT3 4.0: 104 variables, 4 equality constraints
   For improved efficiency, SDPT3 is solving the dual problem.
------------------------------------------------------------

 num. of constraints =  4
 dim. of socp   var  =  4,   num. of socp blk  =  1
 dim. of linear var  = 100
*******************************************************************
   SDPT3: Infeasible path-following algorithms
*******************************************************************
 version  predcorr  gam  expon  scale_data
    NT      1      0.000   1        0
it pstep dstep pinfeas dinfeas  gap      prim-obj      dual-obj    cputime
-------------------------------------------------------------------
 0|0.000|0.000|1.6e+03|2.3e+01|4.9e+04|-2.000000e+03  0.000000e+00| 0:0:00| chol  1  1
 1|0.523|0.525|7.6e+02|1.1e+01|3.3e+04|-1.174305e+03 -7.513991e+01| 0:0:00| chol  1  1
 2|0.366|0.386|4.8e+02|6.8e+00|2.7e+04|-6.430873e+02 -2.287892e+02| 0:0:00| chol  1  1
 3|0.247|0.294|3.6e+02|4.8e+00|2.4e+04|-2.901923e+02 -4.170541e+02| 0:0:00| chol  1  1
 4|0.164|0.240|3.0e+02|3.6e+00|2.3e+04|-1.945426e+01 -6.129483e+02| 0:0:00| chol  1  1
 5|0.109|0.224|2.7e+02|2.8e+00|2.3e+04| 2.095194e+02 -8.126945e+02| 0:0:00| chol  1  1
 6|0.084|0.288|2.5e+02|2.0e+00|2.4e+04| 4.355619e+02 -1.044117e+03| 0:0:00| chol  1  1
 7|0.129|0.943|2.2e+02|1.1e-01|2.3e+04| 7.382631e+02 -1.392626e+03| 0:0:00| chol  1  1
 8|0.122|1.000|1.9e+02|2.2e-06|2.7e+04| 1.327483e+03 -2.316356e+03| 0:0:00| chol  1  1
 9|0.720|0.346|5.3e+01|1.7e-06|1.2e+04| 2.505271e+03 -2.620171e+03| 0:0:00| chol  1  1
10|0.749|1.000|1.3e+01|2.0e-07|9.5e+03| 4.258390e+03 -3.243073e+03| 0:0:00| chol  1  1
11|1.000|1.000|6.4e-13|9.9e-08|5.1e+03| 1.695448e+03 -3.379264e+03| 0:0:00| chol  1  1
12|1.000|0.962|5.9e-12|3.2e-08|2.2e+03| 1.483001e+03 -6.791601e+02| 0:0:00| chol  1  1
13|0.963|1.000|1.9e-12|9.0e-09|7.2e+02| 2.406002e+02 -4.755744e+02| 0:0:00| chol  1  1
14|1.000|1.000|2.6e-12|9.0e-10|2.4e+02| 1.396521e+02 -9.596010e+01| 0:0:00| chol  1  1
15|1.000|1.000|2.5e-13|9.1e-11|8.5e+01| 2.447638e+01 -6.007492e+01| 0:0:00| chol  1  1
```

```
16|1.000|0.999|5.9e-14|1.0e-11|2.9e+01| 1.430668e+01 -1.444949e+01| 0:0:00| chol  1  1
17|1.000|1.000|1.1e-14|1.9e-12|1.0e+01| 2.196689e-01 -1.016041e+01| 0:0:00| chol  1  1
18|0.754|0.822|2.6e-14|1.4e-12|6.7e+00|-3.778618e+00 -1.049978e+01| 0:0:00| chol  1  1
19|1.000|0.839|7.7e-13|1.2e-12|3.5e+00|-4.119932e+00 -7.627466e+00| 0:0:00| chol  1  1
20|0.942|0.958|4.2e-14|1.1e-12|2.4e-01|-6.797549e+00 -7.040376e+00| 0:0:00| chol  1  1
21|0.985|0.974|6.3e-14|1.0e-12|4.1e-03|-6.996935e+00 -7.000998e+00| 0:0:00| chol  1  1
22|0.987|0.982|5.7e-14|1.0e-12|5.6e-05|-6.999961e+00 -7.000017e+00| 0:0:00| chol  1  1
23|0.989|0.985|1.6e-13|1.0e-12|1.5e-06|-6.999999e+00 -7.000000e+00| 0:0:00| chol  1  1
24|0.989|0.987|2.0e-13|1.0e-12|3.8e-08|-7.000000e+00 -7.000000e+00| 0:0:00|
  stop: max(relative gap, infeasibilities) < 1.49e-08
-------------------------------------------------------------------
 number of iterations   = 24
 primal objective value = -6.99999997e+00
 dual   objective value = -7.00000000e+00
 gap := trace(XZ)       = 3.79e-08
 relative gap           = 2.53e-09
 actual relative gap    = 2.53e-09
 rel. primal infeas (scaled problem)   = 1.99e-13
 rel. dual     "        "       "      = 1.01e-12
 rel. primal infeas (unscaled problem) = 0.00e+00
 rel. dual     "        "       "      = 0.00e+00
 norm(X), norm(y), norm(Z) = 1.2e+01, 1.7e+01, 4.9e+01
 norm(A), norm(b), norm(C) = 3.5e+01, 2.0e+00, 1.1e+01
 Total CPU time (secs)  = 0.12
 CPU time per iteration = 0.00
 termination code       =  0
 DIMACS: 2.0e-13  0.0e+00  5.6e-12  0.0e+00  2.5e-09  2.5e-09
-------------------------------------------------------------------

------------------------------------------------------------
Status: Solved
Optimal value (cvx_optval): +6.5
```

```
w
```

```
w = 2×1
    3.0000
    2.0000
```

```
b
```

```
b = -15.0000
```

```
xv = (2:0.01:4);
plot(xv, (15 - 3 * xv) / 2);
```

(c)

```
P = zeros(100, 100);
for i = 1:100
    for j = 1:100
        P(i,j) = y(i) * y(j) * x(i,:) * x(j,:)';
    end
end
cvx_begin
    variable a(100)
    maximize( sum(a) - 1/2 * quad_form(a, P) )
    subject to
        sum(a .* y) == 0
        for i = 1:100
```

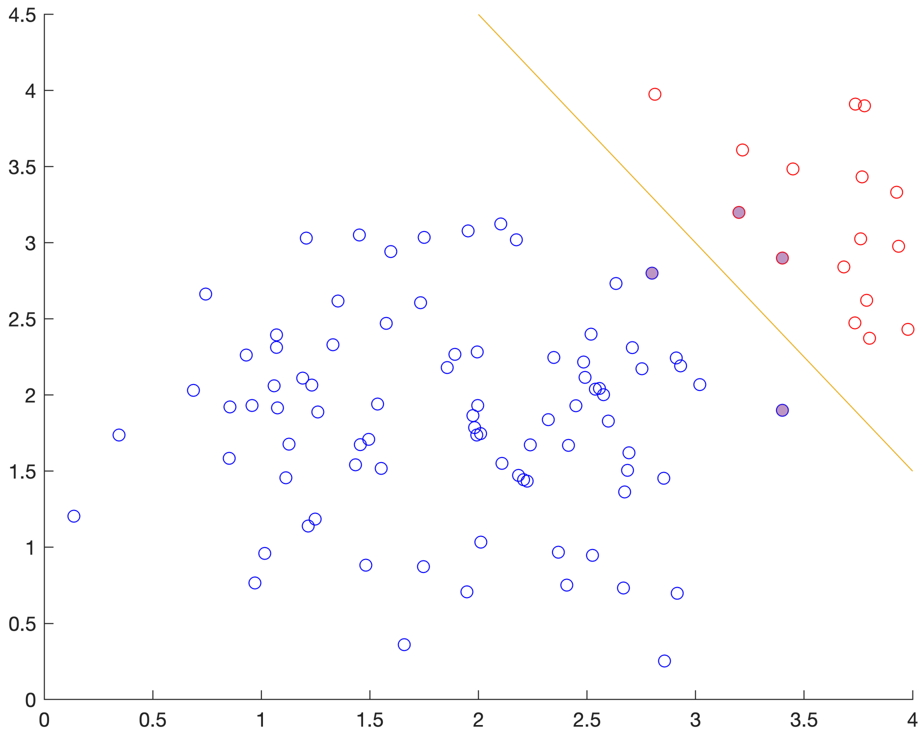
```
            a(i) >= 0
        end
cvx_end
```

```
Calling SDPT3 4.0: 104 variables, 4 equality constraints
------------------------------------------------------------

 num. of constraints =  4
 dim. of socp   var  =  4,   num. of socp blk  =  1
 dim. of linear var  = 100
*******************************************************************
   SDPT3: Infeasible path-following algorithms
*******************************************************************
 version  predcorr  gam  expon  scale_data
    NT      1      0.000   1        0
it pstep dstep pinfeas dinfeas  gap      prim-obj      dual-obj    cputime
-------------------------------------------------------------------
 0|0.000|0.000|6.8e+02|8.3e-01|2.3e+04|-1.480616e+03  0.000000e+00| 0:0:00| chol  1  1
 1|0.750|0.996|1.7e+02|4.9e-03|6.8e+03|-4.713686e+02 -9.722811e+01| 0:0:00| chol  1  1
 2|0.972|1.000|4.8e+00|2.7e-04|2.7e+02| 2.188681e+01 -5.537854e+01| 0:0:00| chol  1  1
 3|0.969|1.000|1.5e-01|2.7e-05|1.0e+02| 4.551568e+01 -4.852598e+01| 0:0:00| chol  1  1
 4|1.000|1.000|3.1e-07|2.7e-06|5.0e+01| 1.599313e+01 -3.367612e+01| 0:0:00| chol  1  1
 5|1.000|1.000|2.6e-08|3.3e-07|1.6e+01| 5.895252e+00 -9.707110e+00| 0:0:00| chol  1  1
 6|1.000|0.593|2.6e-09|1.6e-07|7.3e+00|-1.614783e+00 -8.923866e+00| 0:0:00| chol  1  1
 7|0.918|1.000|1.2e-08|3.3e-09|2.4e+00|-4.590305e+00 -7.007706e+00| 0:0:00| chol  1  1
 8|0.981|0.974|2.8e-09|1.1e-09|7.9e-02|-6.440585e+00 -6.520081e+00| 0:0:00| chol  1  1
 9|0.988|0.987|1.1e-09|5.9e-10|9.6e-04|-6.499293e+00 -6.500255e+00| 0:0:00| chol  1  1
10|0.988|0.981|1.4e-11|2.4e-10|1.3e-05|-6.499991e+00 -6.500004e+00| 0:0:00| chol  1  1
11|0.997|0.954|1.7e-14|1.4e-11|6.5e-07|-6.500000e+00 -6.500000e+00| 0:0:00| chol  1  1
12|1.000|0.924|1.0e-14|2.0e-12|4.8e-08|-6.500000e+00 -6.500000e+00| 0:0:00|
  stop: max(relative gap, infeasibilities) < 1.49e-08
-------------------------------------------------------------------
 number of iterations   = 12
 primal objective value = -6.49999997e+00
 dual   objective value = -6.50000001e+00
 gap := trace(XZ)       = 4.84e-08
 relative gap           = 3.46e-09
 actual relative gap    = 3.39e-09
 rel. primal infeas (scaled problem)   = 1.03e-14
 rel. dual     "        "       "      = 2.03e-12
 rel. primal infeas (unscaled problem) = 0.00e+00
 rel. dual     "        "       "      = 0.00e+00
 norm(X), norm(y), norm(Z) = 7.8e+00, 8.4e+01, 3.8e+02
 norm(A), norm(b), norm(C) = 1.1e+01, 2.0e+00, 3.7e+02
 Total CPU time (secs)  = 0.08
 CPU time per iteration = 0.01
 termination code       =  0
 DIMACS: 1.0e-14  0.0e+00  2.9e-12  0.0e+00  3.4e-09  3.5e-09
-------------------------------------------------------------------

------------------------------------------------------------
Status: Solved
Optimal value (cvx_optval): +6.5
```

```
scatter(GPA(a >= 0.00001), GRE(a >= 0.00001), 'filled', 'MarkerFaceAlpha', 0.5);
```


4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0
0.5
1
1.5
2
2.5
3
3.5
4

4. $K = \begin{bmatrix} k(x_1,x_1) & k(x_1,x_2) \\ k(x_2,x_1) & k(x_2,x_2) \end{bmatrix}$ $\quad y = \begin{bmatrix} k(x_2,x_2) \\ -k(x_1,x_2) \end{bmatrix}$

$$y^T K y = \begin{bmatrix} k(x_2,x_2) & -k(x_1,x_2) \end{bmatrix} \begin{bmatrix} k(x_1,x_1) & k(x_1,x_2) \\ k(x_2,x_1) & k(x_2,x_2) \end{bmatrix} \begin{bmatrix} k(x_2,x_2) \\ -k(x_1,x_2) \end{bmatrix}$$

$$= \big[k(x_2,x_2)k(x_1,x_1) - k(x_1,x_2)k(x_1,x_2),$$
$$k(x_2,x_2)k(x_1,x_2) - k(x_1,x_2)k(x_2,x_2) \big] \begin{bmatrix} k(x_2,x_2) \\ -k(x_1,x_2) \end{bmatrix}$$

$$= k(x_2,x_2)k(x_1,x_1)k(x_2,x_2) - k(x_1,x_2)k(x_1,x_2)k(x_2,x_2)$$

$$\cancel{- k(x_2,x_2)k(x_1,x_2)k(x_1,x_2) + k(x_1,x_2)k(x_2,x_2)k(x_1,x_2)} \to 0$$

$$\cancel{- k(2,2)k(1,2)^2} + k \quad = k(x_2,x_2)\Big(k(x_1,x_1)k(x_2,x_2) - k(x_1,x_2)^2\Big) \geq 0$$

$k(x_1,x_2) = k(x_2,x_1)$
Symmetry.

≥ 0

PSD therefore,

$$k(x_1,x_1)k(x_2,x_2) - k(x_1,x_2)^2 \geq 0$$

$$k(x_1,x_2)^2 \leq k(x_1,x_1)k(x_2,x_2)$$

2

www.septemberleather.com

5. (a) PSD states that K is valid if $y^T K y \geq 0 \quad \forall y$

$$y^T K y = y^T \left(K_1(x,x') + K_2(x,x') \right) y$$

$$= \left(y^T K_1(x,x') + y^T K_2(x,x') \right) y$$

$$= y^T K_1(x,x') y + y^T K_2(x,x') y$$

each ≥ 0 so total sum ≥ 0 ▫

(b) ~~$y^T (K) y = y^T (K_1 K_2) y$~~

~~If K_1 is a valid kernel, K_1 is PSD.~~

~~$y^T K_1 y \geq 0$~~

~~Let $y = v^T$~~

~~$v K_1 v^T \geq 0$~~

$$K_1(x,x') = a(x)^T a(x')$$

$$K_2(x,x') = b(x)^T b(x')$$

$$K = K_1 K_2 = \sum_{m=1}^{M} a_m(x) a_m(x') \cdot \sum_{m=1}^{N} b_m(x) b_m(x')$$

$$= \sum_{m=1}^{M} \sum_{n=1}^{N} a_m(x) b_n(x) \left(a_m(x') b_n(x') \right)$$

Let $C_{m,n}(x) = a_m(x) \cdot b_n x$

$$= \sum_{m=1}^{M} \sum_{n=1}^{N} C_{mn}(x) \cdot C_{mn}(x')$$

C is $m \times n$ dimensional

$$= C(x)^T C(x')$$

K can be written as a inner product of ~~a~~ feature map C.
so K is a kernel ▫.

(c) $\exp(K_1(x,x')) \Rightarrow \exp(a) = 1 + a + \frac{a^2}{2} + \frac{a^3}{6} + \ldots$

$$1 + K_1(x,x') + \frac{1}{2} K_1(x,x')^2 + \frac{1}{6} K_1(x,x')^3 + \ldots$$

each of these terms $K_1(x,x')^n \cdot \frac{1}{n!}$ is a kernel by part (b).
the sum of all of these kernels is a kernel by part (a). ▫

6.

$$t_n(w^T \cdot x_n + b) = 1 - \varepsilon$$

$$w^T x_n + b = t_n(1-\varepsilon)$$

$$\left(\sum_{k \in S} a_k t_k x_k\right)^T x_n + b = t_n(1-\varepsilon)$$

0/M includes all indices that $y^{(i)}(w^T x^{(i)} + b) \geq 1 - \varepsilon_i$

$$|M| \cdot b = \sum_{n \in M}\left(t_n \boxed{(1-\varepsilon)} - \left(\sum_{k \in S} a_k t_k x_k\right)^T x_n\right)$$

$$N_M b = \sum_{n \in M}\left(t_n - \left(\sum_{k \in S} a_k t_k x_k\right)^T x_n\right)$$

$$= \sum_{n \in M}\left(y^{(n)} - \sum_{k \in S} a_k \cdot y^{(k)} \cdot x_k^T \cdot x_n\right)$$

$$= \sum_{n \in M}\left(y^{(n)} - \sum_{m \in S} a_m \cdot y^{(n)} \langle x^{(n)}, x^{(m)} \rangle\right)$$

$$b = \frac{1}{N_M} \sum_{n \in M}\left(y^{(n)} - \sum_{m \in S} a_m \cdot y^{(n)} \langle x^{(n)}, x^{(m)} \rangle\right)$$