204079453
205714447

Homework #5 - Theory + SVM

1. Kernels and mapping functions (25 pts)
   a. (20 pts) Let $K(x,y) = (x \cdot y + 1)^3$ be a function over $\mathbb{R}^2 \times \mathbb{R}^2$ (i.e., $x, y \in \mathbb{R}^2$).

      Find ψ for which K is a kernel. (It may help to first expand the above term on the right-hand side).
   b. (2 pts) What did we call the function ψ in class if we remove all coefficients?
   c. (3 pts) How many multiplication operations do we save by using $K(x,y)$ versus $\psi(x) \cdot \psi(y)$?

* $(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$

♡ $x \cdot y = (x_1y_1 + x_2y_2)$

a. $(x\cdot y+1)^{3*} = (x\cdot y)^3 + 3(x\cdot y)^2 + 3x\cdot y + 1$

* $(x_1y_1+x_2y_2)^3 = (x_1^3y_1^3 + 3x_1^2y_1^2x_2y_2 + 3x_1y_1x_2^2y_2^2 + x_2^3y_2^3)$

* $3(x_1y_1+x_2y_2)^2 = 3[x_1^2y_1^2 + 2x_1y_1x_2y_2 + x_2^2y_2^2]$

* $3(x_1y_1+x_2y_2)$

↓ plug in the equation:

$$= (x_1^3y_1^3 + 3x_1^2y_1^2x_2y_2 + 3x_1y_1x_2^2y_2^2 + x_2^3y_2^3) + 3[x_1^2y_1^2 + 2x_1y_1x_2y_2 + x_2^2y_2^2] + 3(x_1y_1+x_2y_2) + 1$$

$$= (x_1^3, \sqrt{3}x_1^2x_2, \sqrt{3}x_1x_2^2, x_2^3, \sqrt{3}x_1^2, \sqrt{6}x_1x_2, \sqrt{3}x_2^2, \sqrt{3}x_1, \sqrt{3}x_2, 1) \cdot$$
$$(y_1^3, \sqrt{3}y_1^2y_2, \sqrt{3}y_1y_2^2, y_2^3, \sqrt{3}y_1^2, \sqrt{6}y_1y_2, \sqrt{3}y_2^2, \sqrt{3}y_1, \sqrt{3}y_2, 1)$$

therefore: $\varphi(x) = (x_1^3, \sqrt{3}x_1^2x_2, \sqrt{3}x_1x_2^2, x_2^3, \sqrt{3}x_1^2, \sqrt{6}x_1x_2, \sqrt{3}x_2^2, \sqrt{3}x_1, \sqrt{3}x_2, 1)$

b. Removing all the coeficents will generate a full rational variety of order 3 & dimension 2.

c. # of multiplication operation for $\varphi(x)\cdot\varphi(y) = 10$ (10 indeces, $\binom{5}{3} = 10$)

# of multiplication operation for $K(x,y) = 4$

therefore with the kernel we saved $(10-4) = 6$ multiplication operation

2. Lagrange multipliers (25 pts)

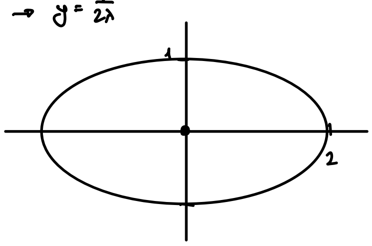
Let $f(x,y) = 2x - y$. Find the minimum and the maximum points for f under the constraint $g(x,y) = \frac{x^2}{4} + y^2 = 1$.

2) $L(x,y,\lambda) = 2x - y + \lambda\left(\frac{x^2}{4} + y^2 - 1\right)$

* $\frac{\partial L}{\partial x} = 2 + \lambda \cdot \frac{2x}{4} = 0 \to 2 + \frac{1}{2}\lambda x = 0 \;/\cdot 2 \to 4 + \lambda x = 0 \to \lambda x = -4 \;/:\lambda,\ \lambda \neq 0 \to x = -\frac{4}{\lambda}$

* $\frac{\partial L}{\partial y} = -1 + \lambda \cdot 2y = 0 \;/:2 \to -\frac{1}{2} + \lambda y = 0 \to \lambda y = \frac{1}{2} \;/:\lambda,\ \lambda \neq 0 \to y = \frac{1}{2\lambda}$

* $\frac{\partial L}{\partial \lambda} = g(x,y) = \frac{x^2}{4} + y^2 - 1 = 0$ (elipse as drawn: $\to$



$\underline{x = -\frac{4}{\lambda},\ y = \frac{1}{2\lambda},\ \lambda \neq 0:}$ $\quad \frac{x^2}{4} + y^2 - 1 = 0$

$\frac{\left(-\frac{4}{\lambda}\right)^2}{4} + \left(\frac{1}{2\lambda}\right)^2 - 1 = 0$

$\frac{\frac{16}{\lambda^2}}{4} + \frac{1}{4\lambda^2} - 1 = 0$

$\frac{4}{\lambda^2} + \frac{1}{4\lambda^2} - 1 = 0 \;/\cdot 4\lambda^2,\ \lambda \neq 0$

$16 + 1 - 4\lambda^2 = 0 \to 4\lambda^2 = 17 \to \lambda^2 = \frac{17}{4} \to \lambda = \pm\frac{\sqrt{17}}{2}$

\#1

$\boxed{\lambda_1 = \frac{\sqrt{17}}{2}}$: $\quad x_1 = -\frac{4}{\lambda} \to \frac{-4}{\frac{\sqrt{17}}{2}} \to -\frac{8}{\sqrt{17}} \rightsquigarrow \boxed{x_1 = -\frac{8}{\sqrt{17}}}$

$y_1 = \frac{1}{2\lambda} \to \frac{1}{\cancel{2} \cdot \frac{\sqrt{17}}{\cancel{2}}} \rightsquigarrow \boxed{y_1 = \frac{1}{\sqrt{17}}}$

\#2

$\boxed{\lambda_2 = -\frac{\sqrt{17}}{2}}$: $\quad x_2 = -\frac{4}{\lambda} \to \frac{-4}{-\frac{\sqrt{17}}{2}} \to \frac{8}{\sqrt{17}} \rightsquigarrow \boxed{x_2 = \frac{8}{\sqrt{17}}}$

$y_2 = \frac{1}{2\lambda} \to \frac{1}{\cancel{2} \cdot \left(-\frac{\sqrt{17}}{\cancel{2}}\right)} \rightsquigarrow \boxed{y_2 = \frac{-1}{\sqrt{17}}}$

$f(x_1,y_1) = 2\cdot\left(-\frac{8}{\sqrt{17}}\right) - \frac{1}{\sqrt{17}} = \frac{-17}{\sqrt{17}} = -\sqrt{17}$

$f(x_2,y_2) = 2\left(\frac{8}{\sqrt{17}}\right) - \left(-\frac{1}{\sqrt{17}}\right) = \frac{17}{\sqrt{17}} = \sqrt{17}$

therefore: $f(x_1,y_1) < f(x_2,y_2)$

$\downarrow$

$(x_1,y_1) = \left(-\frac{8}{\sqrt{17}}, \frac{1}{\sqrt{17}}\right)$ is a minimum point for f under the constraint $g(x,y) = \frac{x^2}{4} + y^2 = 1$

with $\lambda = \frac{\sqrt{17}}{2}$

$(x_2,y_2) = \left(\frac{8}{\sqrt{17}}, -\frac{1}{\sqrt{17}}\right)$ is a maximum point for f under the constraint $g(x,y) = \frac{x^2}{4} + y^2 = 1$

with $\lambda = -\frac{\sqrt{17}}{2}$

3. PAC Learning (25 pts)

Let $X = \mathbb{R}^2$. Let vectors $u = \left(\frac{\sqrt{3}}{2}, \frac{1}{2}\right), \overset{*}{w} = \left(\frac{\sqrt{3}}{2}, -\frac{1}{2}\right), v = (0, -1)$

and $C = H = \left\{ h(r) = \left\{ (x_1, x_2) \middle| \begin{array}{l} (x, y) \cdot u \le r, \\ (x, y) \cdot v \le r, \\ (x, y) \cdot w \le r \end{array} \right\} \right\}$, for $r > 0$,

the set of all origin-centered upright equilateral triangles.

Describe a polynomial sample complexity algorithm L that learns C using H. State the time complexity and the sample complexity of your suggested algorithm. Prove all your steps.

from the Piazza:
$* \; w = \left(-\frac{\sqrt{3}}{2}, \frac{1}{2}\right)$

Algoritm $L(D)$:

$D \in X^m$, $\forall d \in D \rightarrow d = (x_1, x_2)$, $x_1, x_2 \in \mathbb{R}$

algo:

$r = 0$ / initialize r with a number close to zero since $r > 0$.

for every d in D:

$r_{max} = \arg\max \{ d \cdot u, d \cdot v, d \cdot w \}$

if $r_{max} > \delta$:

$r = r_{max}$

return $h(r)$, $h \in H$.

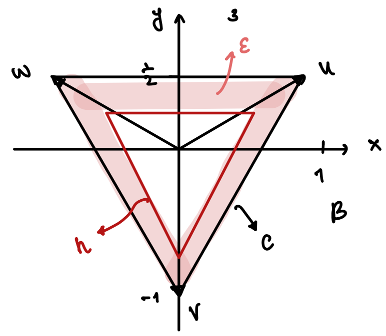
algo Complexity $O(m)$

Time Complexity - since we iterate over m samples and make operations with $O(1)$ therefore the total time complexity will be $O(m)$ where m is polynomial in $1/\varepsilon$ & $1/\delta$

(Continue to the sample complexity..)

Sample Complexity:

* δ - Confidence level $0 \le \delta \le 1$



$p(err > \varepsilon) \le \delta$

$p(\{D \in X^m : Err(L(D), c) > \varepsilon\}) \le$

$p(X-B)^m \le (1-\varepsilon)^m \le e^{-m\varepsilon}$

* $e^{-m\varepsilon} \le \delta$

$e^{-m\varepsilon} \le \delta$

$-m\varepsilon \le \ln(\delta) / \cdot(-1)$

$m\varepsilon \ge -\ln(\delta) / \frac{1}{\varepsilon}$

$$\boxed{m \ge \frac{1}{\varepsilon} \cdot \left[-\ln(\delta) \right]}$$

For example, with certinty of 95% $(1-\delta)$ and err of 10% (ε)

$m \ge \frac{1}{0.1}\left[-\ln(0.05)\right] \rightarrow m \gtrsim 30$

therefore the sample size needs to be at least 30 .

4. (15 pts) A business manager at your ecommerce company asked you to make a model to predict whether a user is going to proceed to checkout or abandon their cart. You created the model using, and reported 20% error on your test set of size 1000 samples. In the business manager's presentation to upper management, he presented your model and stated that the company can expect 20% error when deploying the model live on the website.

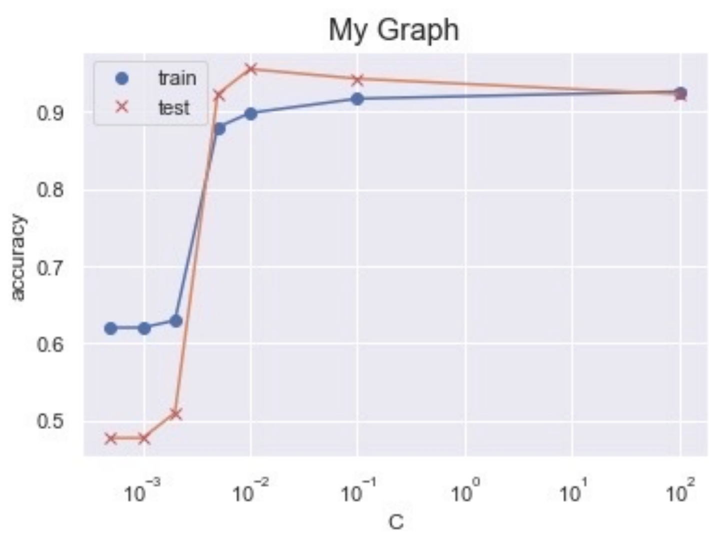
Luckily, you realize that this is a mistaken assumption, and you correct the statement to say that with 95% confidence, the true error they can expect is up to what percentage? (Just state the error percentage).

upper bound: $\hat{p} + 2SE = \hat{p} + 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.2 + 2\sqrt{\frac{0.2 \cdot 0.8}{1000}} = 0.2 + 2 \cdot 0.0126 = 0.225 = \boxed{22.5\%}$

5. SVM (10 pts)

See the notebook in the homework files and follow the instructions there.

Take a **screenshot** of your resulting graph near the bottom of the notebook (titled "My Graph") and paste into your submission PDF along with your answers to the theoretical questions. Do **NOT** submit your code.




```
Cs = [0.0005, 0.001, 0.002, 0.005, 0.01, 0.1, 100]
```