HW5 – Theory + SVM

1. <u>PAC Learning and VC dimension (30 pts)</u>

   Let $X = \mathbb{R}^2$. Let

   $C = H = \left\{ h(r_1, r_2) = \left\{ (x_1, x_2) \middle| \begin{matrix} x_1^2 + x_2^2 \geq r_1 \\ x_1^2 + x_2^2 \leq r_2 \end{matrix} \right\} \right\}$, for $0 \leq r_1 \leq r_2$,

   the set of all origin-centered rings.

   a. (8 pts) What is the $VC(H)$? Prove your answer.
   b. (14 pts) 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.

   In class we saw a bound on the sample complexity when H is finite.

   $$m \geq \frac{1}{\varepsilon}\left(\ln|H| + \ln\frac{1}{\delta}\right)$$

   When $|H|$ is infinite, we have a different bound:

   $$m \geq \frac{1}{\varepsilon}\left(4\log_2\frac{2}{\delta} + 8VC(H)\log_2\frac{13}{\varepsilon}\right)$$

   c. (8 pts) You want to get with 95% confidence a hypothesis with at most 5% error. Calculate the sample complexity with the bound that you found in b and the above bound for infinite $|H|$. In which one did you get a smaller m? Explain.

Handwritten answer in next page

2. <u>VC dimension (20 pts)</u>

Let $X = \mathbb{R}$ and $n \in \mathbb{N}$.

Define "x-node decision tree" for any $x = 2^n - 1$ to be a full binary decision tree with x nodes (including the leaves).

Let H_m be the hypothesis space of all "x-node decision tree" with $n \leq m$.

a. (5 pts) What is the $VC(H_3)$? Prove your answer.
b. (15 pts) What is the $VC(H_m)$? Prove your answer.

Handwritten answer in next page

3. 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).

      Answer:

$$Let\ X = (x_1, x_2)\ and\ Y = (y_1, y_2) => \quad X * Y = (x_1 * y_1 + x_2 * y_2)$$

      We'll use the following binomial theorem formulas to expand Kernel function:

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

$$\begin{aligned} K(x, y) = (x \cdot y + 1)^3 = \left((x_1y_1 + x_2y_2) + 1\right)^3 \\ &= (x_1 * y_1 + x_2 * y_2)^3 + 3(x_1 * y_1 + x_2 * y_2)^2 \\ &+ 3(x_1 * y_1 + x_2 * y_2) + 1 \\ &= \left((x_1^3y_1^3 + 3x_1^2y_1^2x_2y_2 + 3x_1y_1x_2^2y_2^2 + x_2^3y_2^3)\right. \\ &\left.+ 3(x_1^2y_1^2 + 2x_1y_1x_2 + x_2^2y_2^2) + 3x_1y_1 + 3x_2y_2 + 1\right) \end{aligned}$$

      Each term in this expansion represents one element in the feature mapping $\psi(\mathrm{x})$, So K is a kernel of:

$$\boldsymbol{\psi(\mathrm{x}) = (x_1^3, \sqrt{3}x_1^2x_2, \sqrt{3}x_1x_2^2, x_2^3, \sqrt{3}x_1^2, \sqrt{6}x_1x_2, \sqrt{3}x_2^2, \sqrt{3}x_1, \sqrt{3}x_2, 1)}$$

   b. (2 pts) What did we call the function ψ in class if we remove all coefficients?

   Answer:

   If we remove all coefficients we will have:

$$\psi(\mathrm{x}) = (x_1^3, x_1^2x_2, x_1x_2^2, x_2^3, x_1^2, x_1x_2, x_2^2, x_1, x_2, 1)$$

   We called this a full rational variety of order 3

   This is in an input space of dimension 2 is described by all 3-th degree monomials of the input variables in x.

   The size of this group should be $\genfrac{}{}{0pt}{}{r+n}{r}$ while here $r = 3$ and $n = 2$ so we have $\genfrac{}{}{0pt}{}{3+2}{3} = \genfrac{}{}{0pt}{}{5}{3} = 10$ items here.

c. (3 pts) How many multiplication operations do we save by using $K(x, y)$ versus $\psi(x) \cdot \psi(y)$?

Answer:

Using K, we will have 2 multiplications for the dot product over $\mathbb{R}^2 \times \mathbb{R}^2$

Using $\psi(x) \cdot \psi(y)$, we will have 10 multiplications for the dot product of 10 elements in the feature space.

Therefore, we will save $10 - 2 = 8$ multiplication operations by using the Kernel function.

4. <u>Lagrange multipliers (15 pts)</u>

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

Answer:

To find the minimum and maximum points using Lagrange multipliers, we define the Lagrangian function:

$$L(x, y, \lambda) = f(x, y) + \lambda(g(x, y) - 1)$$

$$f(x, y) = 2x - y \; and \; g(x, y) = \frac{x^2}{4} + y^2.$$

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

$$\frac{\partial L}{\partial x} = 2 + \lambda(\frac{x}{2}) = 0$$
$$\frac{\partial L}{\partial y} = -1 + 2\lambda y = 0$$
$$\frac{\partial L}{\partial \lambda} = \frac{x^2}{4} + y^2 - 1 = 0$$

$$From \frac{\partial L}{\partial x} = 0:$$
$$2 + \lambda(\frac{x}{2}) = 0$$
$$\lambda x = -4$$
$$\lambda = -\frac{4}{x}$$

$$From \frac{\partial L}{\partial y} = 0:$$
$$-1 + 2\lambda y = 0$$
$$\lambda y = \frac{1}{2}$$
$$y = \frac{1}{2\lambda}$$

$$Substitute \; \lambda = -\frac{4}{x} \; into \; y = \frac{1}{2\lambda}:$$
$$y = \frac{1}{2\left(-\frac{4}{x}\right)}$$
$$y = -\frac{x}{8}$$
$$Substitute \; y = -\frac{x}{8} \; into \; the \; constraint \frac{x^2}{4} + y^2 = 1:$$

$$\frac{x^2}{4} + \left(-\frac{x}{8}\right)^2 = 1$$
$$\frac{x^2}{4} + \frac{x^2}{64} = 1$$
$$16x^2 + x^2 = 64$$
$$17x^2 = 64$$
$$x^2 = \frac{64}{17}$$
$$x = \pm\left(\frac{8}{\sqrt{17}}\right)$$

$$y = -\frac{x}{8} = \pm\left(-\frac{1}{\sqrt{17}}\right)$$

The critical points are:

$$\left(\frac{8}{\sqrt{17}}, -\frac{1}{\sqrt{17}}\right) \; and \; \left(-\frac{8}{\sqrt{17}}, \frac{1}{\sqrt{17}}\right)$$

Evaluate f(x, y) at these points:

$$f\left(\frac{8}{\sqrt{17}}, -\frac{1}{\sqrt{17}}\right) = 2\left(\frac{8}{\sqrt{17}}\right) - \left(-\frac{1}{\sqrt{17}}\right) = \left(\frac{16}{\sqrt{17}}\right) + \left(\frac{1}{\sqrt{17}}\right) = \frac{17}{\sqrt{17}} = \sqrt{17}$$
$$f\left(-\frac{8}{\sqrt{17}}, \frac{1}{\sqrt{17}}\right) = 2\left(-\frac{8}{\sqrt{17}}\right) - \left(\frac{1}{\sqrt{17}}\right) = \left(-\frac{16}{\sqrt{17}}\right) - \left(\frac{1}{\sqrt{17}}\right) = -\frac{17}{\sqrt{17}} = -\sqrt{17}$$

So, our finals results are:

$$\boldsymbol{The\ maximum\ value\ of\ f(x,y)\ is\ \sqrt{17}\ at\ the\ point\ \left(\frac{8}{\sqrt{17}}, -\frac{1}{\sqrt{17}}\right).}$$
$$\boldsymbol{The\ minimum\ value\ of\ f(x,y)\ is\ -\sqrt{17}\ at\ the\ point\ \left(-\frac{8}{\sqrt{17}}, \frac{1}{\sqrt{17}}\right).}$$

5. See notebook exercise (10 pts)

Answer in the attached Jupyter notebook