Please provide complete and well-written solutions to the following exercises.

Due October 11, at the beginning of class.

Homework 2

Exercise 1. Let M be a $k\times k$ real symmetric matrix. Then M is positive semidefinite if and only if there exists a real $k\times k$ matrix R such that

$$M = RR^T.$$

In either case, if $r^{(i)}$ denotes the i^{th} row of R, we have

$$m_{ij} = \langle r^{(i)}, r^{(j)}\rangle, \qquad \forall\, 1\leq i,j\leq k.$$

Exercise 2. Let μ be a Borel measure on $\mathbf{R}^n$ such that the measure of any open set in $\mathbf{R}^n$ is positive. Let $m\colon \mathbf{R}^n\times\mathbf{R}^n\to\mathbf{R}$ be continuous with $\int_{\mathbf{R}^n}\int_{\mathbf{R}^n}\left|m(x,y)\right|^2 d\mu(x)d\mu(y)<\infty$. Show that the following two positive semidefinite conditions on m are equivalent:

- $\forall\ p\geq 1$, for all $z^{(1)},\ldots,z^{(p)}\in\mathbf{R}^n$, for all $\beta_1,\ldots,\beta_p\in\mathbf{R}$ we have
$$\sum_{i,j=1}^{p}\beta_i\beta_j m(z^{(i)},z^{(j)})\geq 0.$$
- $\forall\ f\in L_2(\mu)$, we have
$$\int_{\mathbf{R}^n}\int_{\mathbf{R}^n} f(x)f(y)m(x,y)d\mu(x)d\mu(y)\geq 0.$$

From either condition, we should see that the converse of Mercer's Theorem holds. We should also be able to deduce various properties of positive semidefinite (PSD) kernels. For example, a nonnegative linear combination of PSD kernels is PSD.

Exercise 3. For each kernel function $m\colon\mathbf{R}^n\times\mathbf{R}^n\to\mathbf{R}$ below, find an inner product space C and a map $\phi\colon\mathbf{R}^n\to C$ such that

$$m(x,y) = \langle\phi(x),\phi(y)\rangle_C, \qquad \forall\, x,y\in\mathbf{R}^n.$$

Conclude that each such m is a positive semidefinite function, in the sense stated in Mercer's Theorem.

- $m(x,y) := 1+\langle x,y\rangle\ \forall\ x,y\in\mathbf{R}^n$.
- $m(x,y) := (1+\langle x,y\rangle)^d\ \forall\ x,y\in\mathbf{R}^n$, where d is a fixed positive integer.
- $m(x,y) := \exp(-\left\|x-y\right\|^2)$.

Hint: it might be helpful to consider d-fold iterated tensor products of the form $x^{\otimes d} = x\otimes x\otimes\cdots\otimes x$, along with their corresponding inner products.

Exercise 4. Show that the set of conjunctions is contained in the set of linear threshold functions. That is, given a boolean conjunction $f\colon \{0,1\}^n \to \{0,1\}$, find $w \in \mathbf{R}^n, t \in \mathbf{R}$ such that

$$f(x) = 1_{\{\langle w,x\rangle > t\}}, \qquad \forall\, x = (x_1,\ldots,x_n) \in \{0,1\}^n.$$

Exercise 5. Here is an elementary example of "boosting" for random variables.

Suppose X is a real-valued random variable, and $X_1, X_2, \ldots$ are independent copies of X. Let $a < b$, $a, b \in \mathbf{R}$. Suppose it is known that

$$\mathbb{P}(a \leq X \leq b) > 3/4.$$

Fix a positive integer n. Let Y_n be a median of $X_1, \ldots, X_n$. Then Y_n is a "boosted" version of X in the sense that

$$\mathbb{P}(a \leq Y_n \leq b) \geq 1 - \sum_{j=\lfloor n/2 \rfloor}^{n} \binom{n}{j} \alpha^j,$$

where $\alpha := \mathbb{P}(X \notin [a,b])$.

(Optional:) Show additionally that

$$\mathbb{P}(a \leq Y_n \leq b) \geq 1 - (1+o(1))\sqrt{\frac{\pi}{2}}\frac{1}{\sqrt{n}}\frac{2^n \alpha^{\lfloor n/2 \rfloor}}{1-\alpha} \geq 1 - (4\alpha)^{\lfloor n/2 \rfloor} \cdot O(1).$$

Exercise 6. Explain why taking the expected value of the inequality for the average number of mis-classifications of Adaboost does not guarantee PAC learning.

Exercise 7. Show that the Sauer-Shelah lemma is sharp for all n, d. That is, find $\mathcal{F}$ with $d := \mathrm{VCdim}(\mathcal{F})$ such that

$$|\mathcal{F}| = \sum_{i=0}^{d} \binom{n}{i}.$$

(Hint: consider the set of $x \in \{0,1\}^n$ such that x has at most d entries equal to 1.)

Exercise 8. Show that both our notions of ε-net agree (up to changing the constant ε) in the following case: Ω is a metric space, $\mathbb{P}$ is a probability law on Ω, $A = \{B(x,r)\colon x \in \Omega, r > 0\}$ and there exist $a, b, c_1, c_2 > 0$ such that $c_1 r^a \leq \mathbb{P}(B(x,r)) \leq c_2 r^b$ for all $x \in \Omega, r > 0$.

Exercise 9. For any $f \in \mathcal{F}$, show that

$$\mathrm{VCdim}(\mathcal{F}) = \mathrm{VCdim}(D(f)).$$

(Recall: $\mathcal{F}$ is a subset of $\{0,1\}$-valued functions on a set A. Let $f, g \in \mathcal{F}$. Since $f = 1_{\{f=1\}}$, we can identify f with the set where it is 1 and extend set operations to functions in $\mathcal{F}$. For example, $f\Delta g := 1_{\{f=1\}\Delta\{g=1\}}$, where Δ denotes symmetric difference. And we define

$$D(f) := \{f\Delta g\colon g \in \mathcal{F}\}.)$$