**Math Foundations of ML, Fall 2017**

**Homework #4**

**Due Monday September 25, at the beginning of class**

**As stated in the syllabus, unauthorized use of previous semester course materials is strictly prohibited in this course.**

1. Using you class notes, prepare a 1-2 paragraph summary of what we talked about in class in the last week. I do not want just a bulleted list of topics, I want you to use complete sentences and establish context (Why is what we have learned relevant? How does it connect with other things you have learned here or in other classes?). The more insight you give, the better.

2. Let $\mathcal{E}$ be the space of signals on $[-1,1]$ that are even:

$$x(t) \in \mathcal{E} \quad \Leftrightarrow \quad x(t) = x(-t), \quad t \in [-1,1],$$

and let $\mathcal{O}$ be the space of signals on $[-1,1]$ that are odd:

$$x(t) \in \mathcal{O} \quad \Leftrightarrow \quad x(t) = -x(-t), \quad t \in [-1,1].$$

(a) Given an arbitrary $x(t) \in L_2([-1,1])$ what is the closest even function to $\boldsymbol{x}$? That is, solve

$$\min_{\boldsymbol{y} \in \mathcal{E}} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2.$$

A good way to do this is to use the *orthogonality principle* — we know that for the optimal $\hat{\boldsymbol{y}} \in \mathcal{E}$

$$\langle \boldsymbol{x} - \hat{\boldsymbol{y}}, \boldsymbol{z} \rangle = 0 \quad \text{for all} \quad \boldsymbol{z} \in \mathcal{E}.$$

You might consider filling in the blanks in the following line of reasoning:

$$\begin{aligned} \langle \boldsymbol{x} - \hat{\boldsymbol{y}}, \boldsymbol{z} \rangle &= \int_{-1}^{1} [x(t) - y(t)] z(t) \ \mathrm{d}t \\ &= \int_{0}^{1} [x(t) - y(t)] z(t) + \cdots \ \mathrm{d}t \\ &= \cdots \\ &= 0 \quad \text{for all } \boldsymbol{z} \in \mathcal{E} \text{ when } \hat{y}(t) = \cdots . \end{aligned}$$

(b) Given an arbitrary $x(t) \in L_2([-1,1])$, solve

$$\min_{\boldsymbol{y} \in \mathcal{O}} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2.$$

(c) Let $\{\phi_k(t),\ k \geq 0\}$ be an orthobasis for $L_2([0,1])$. How can we use this orthobasis on $[0,1]$ to construct an orthobasis $\{\phi_k^e(t),\ k \geq 0\}$ for $\mathcal{E}$? What about an orthobasis $\{\phi_k^o(t),\ k \geq 0\}$ for $\mathcal{O}$? Is $\{\phi_k^e(t),\ k \geq 0\} \cup \{\phi_k^o(t),\ k \geq 0\}$ an orthobasis for all of $L_2([-1,1])$? Why or why not?

3. (a) The vector space $L_2([0,1]^2)$ is the space of signals of two variables, $x(s,t)$ with $s,t \in [0,1]$ such that
$$\int_0^1 \int_0^1 |x(s,t)|^2 \, \mathrm{d}s \, \mathrm{d}t \; < \; \infty.$$
Let $\{\psi_k(t),\ k \geq 0\}$ be an orthobasis for $L_2([0,1])$. Define
$$v_{k,\ell}(s,t) = \psi_k(s)\,\psi_\ell(t), \quad k,\ell \geq 0.$$
Show that $\{v_{k,\ell}(s,t),\ k,\ell \geq 0\}$ is an orthobasis for $L_2([0,1]^2)$. (You need to argue that the $\boldsymbol{v}_{k,\ell}$ are orthonormal and that they span $L_2([0,1]^2)$.)

(b) Given on orthobasis for $L_2([0,1])$, describe how to construct an orthobasis for $L_2([0,1]^D)$ — the space of functions of $D$ continuous-valued variables $x(\boldsymbol{t})$ such that
$$\int_0^1 \cdots \int_0^1 |x(\boldsymbol{t})|^2 \, \mathrm{d}t_1 \cdots \mathrm{d}t_D \; < \; \infty.$$

4. The Haar basis for the space $L_2([0,1])$ (with the standard inner product) is $\{\boldsymbol{\phi}\} \cup \{\boldsymbol{\psi}_{j,k},\ j \geq 0,\ k = 0,\ldots,2^j - 1\}$, where
$$\phi(t) = \begin{cases} 1, & 0 \leq t \leq 1 \\ 0, & \text{otherwise,} \end{cases} \qquad \psi(t) = \begin{cases} 1, & 0 \leq t \leq 1/2, \\ -1, & 1/2 < t \leq 1, \\ 0, & \text{otherwise,} \end{cases} \qquad \psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k).$$

(a) Sketch the eight functions $\phi(t), \psi_{0,0}(t), \psi_{1,0}(t), \psi_{1,1}(t), \psi_{2,0}(t), \psi_{2,1}(t), \psi_{2,2}(t), \psi_{2,3}(t)$ on separate axes. Give a qualitative description of the 8-dimensional space that they span.

(b) Sketch $\psi_{10,731}(t)$.

(c) What is $\mathcal{S}_J = \operatorname{Span}\left(\{\boldsymbol{\phi},\ \boldsymbol{\psi}_{j,k},\ 0 \leq j \leq J-1,\ \ 0 \leq k \leq 2^j - 1\}\right)$?

(d) A function $f(t)$ is called *Lipschitz* if there exists a constant $C$ such that
$$|f(t_1) - f(t_2)| \; \leq \; C|t_1 - t_2|.$$
All Lipschitz functions are continuous (but they are not necessarily differentiable). Now, suppose $f(t)$ is Lipschitz. Find something reasonable to put on the right-hand side below:
$$|\langle \boldsymbol{f}, \boldsymbol{\psi}_{j,k} \rangle|^2 \; \leq \; \text{(something that depends on } j \text{ and } C).$$
Why does this expression not depend on $k$?

(e) Suppose again that $\boldsymbol{f}$ is Lipschitz, and let
$$\boldsymbol{f}_J = \text{closest point in } \mathcal{S}_J \text{ to } f,$$
where we are again using the standard $L_2$ norm for our notion of "closest". Find something reasonable to put on the right-hand side below:
$$\|\boldsymbol{f} - \boldsymbol{f}_J\|_2 \; \leq \; \text{(something that depends on } j \text{ and } C).$$

5. Let

$$\phi(t) = \left(\frac{2}{\pi}\right)^{1/4} e^{-t^2}, \quad \phi_k(t) = \phi(t-k).$$

These are the same "bell curve bump functions" you considered on the last homework, just renormalized so that $\|\boldsymbol{\phi}_k\|_2 = 1$ and centered slightly differently. Set

$$\mathcal{T}_{10} = \mathrm{Span}\left(\{\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_{10}\}\right).$$

(a) Compute and plot the 10 dual basis vectors $\tilde{\boldsymbol{\phi}}_1, \ldots, \tilde{\boldsymbol{\phi}}_{10}$. These are each functions of a continuous-variable; make your plots on separate axes.

(b) Since $\mathcal{T}_{10}$ is a finite dimensional space, every basis for $\mathcal{T}_{10}$ is a Riesz basis. Show how the Riesz constants $A, B$ (defined on page I.87 of the notes) can be related to the eigenvalues of the Gram matrix for a particular basis, and find $A, B$ for the $\{\boldsymbol{\phi}_k\}_{k=1}^{10}$ defined above.

(c) For the values of $A, B$ computed above, find coefficients $\{c_k\}$ such that

$$\left\|\sum_{k=1}^{10} c_k \boldsymbol{\phi}_k\right\|_2^2 = A \sum_{k=1}^{10} |c_k|^2$$

and $\{d_k\}$ such that

$$\left\|\sum_{k=1}^{10} d_k \boldsymbol{\phi}_k\right\|_2^2 = B \sum_{k=1}^{10} |d_k|^2.$$

Make a plot of the two functions

$$f_c(t) = \sum_{k=1}^{10} c_k \phi_k(t), \quad f_d(t) = \sum_{k=1}^{10} d_k \phi_k(t).$$