MATB24 GRADED PROBLEMS 4, DUE Thursday Nov 12, 11:59am

General instructions:

- You should submit your work on Quercus. The only accepted format is PDF.
- Do not wait until last minute to avoid technical difficulties.
- There is a one point penalty for late submissions within 12 hours of the due date.
- You are encouraged to work in groups, ask questions on Piazza, or in office hours, but you should write your homework individually in your own words. You can get help from me, your TA or your peers, but you should write your solution on your own.
- Unless otherwise stated in all questions you should fully justify your answer.
- Your TA will grade a randomly selected subset of the questions in each homework and your grade will be only based on the graded questions.

Reading assignment:

It is assumed that you read at least one of the reading options below

- Sec 6.1-6.6 from Fraleigh–Beauregard
- Sec 6.A-6.C from Axler

Problem 1. *Let $V_1, V_2, \ldots, V_p$ be mutually orthogonal subspaces of $\mathbb{R}^n$. (In other words, if $i \neq j$, then $\vec{v} \perp \vec{w}$ for all $\vec{v} \in V_i$, $\vec{w} \in V_j$.) Prove that*

$$\dim V_1 + \dim V_2 + \cdots + \dim V_p \leq n.$$

Problem 2. *Let U be an orthogonal $n \times n$ matrix, and consider the linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^n$ defined by $T(\vec{x}) = U\vec{x}$. Let W be a subspace of $\mathbb{R}^n$ such that $T(W) \subseteq W$.*[1]

(a) *Prove that $T(W) = W$.*

(b) *Prove that $T(W^\perp) = W^\perp$.*

Problem 3. *Suppose that V is a real inner product space of dimension n.*

(a) *Show that there exists an isomorphism $T\colon V \to \mathbb{R}^n$ such that*

$$\forall f, g \in V, \qquad \langle f, g \rangle = T(f) \cdot T(g).$$

(b) *Suppose that A is a symmetric invertible $n \times n$ matrix such that*

$$\forall \vec{v} \in \mathbb{R}^n, \qquad \vec{v} \neq \vec{0} \Rightarrow \vec{v}^T A \vec{v} > 0.$$

*Show that there is an invertible $n \times n$ matrix B such that $A = B^T B$. (*Hint: Consider the inner product (you need to check it is an inner product) on $\mathbb{R}^n$ given by $\langle \vec{v}, \vec{w} \rangle = \vec{v}^T A \vec{w}$.*)*

Problem 4. *Suppose that V is a finite-dimensional vector space over $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ with inner product $\langle -, - \rangle\colon V \times V \to \mathbb{F}$.*

(a) *Prove that for every vector $\vec{w} \in V$ the function $V \to \mathbb{F}$ given by $\vec{v} \mapsto \langle \vec{v}, \vec{w} \rangle$ is a linear map.*

(b) *Prove that for every linear map $T\colon V \to \mathbb{F}$ there exists a unique vector $\vec{w} \in V$ such that $T(\vec{v}) = \langle \vec{v}, \vec{w} \rangle$.*

The statement of part (b) is known as the Riesz representation theorem*.*

[1] By definition, $T(W) = \{T(\vec{w}) \mid \vec{w} \in W\}$.

(c) *Recall that $\mathcal{P}_3(\mathbb{R})$ denotes the vector space over $\mathbb{R}$ of polynomials of degree ≤ 3 with real coefficients. Use the Riesz representation theorem to prove that for each $t \in \mathbb{R}$ we can find a unique polynomial $q_t \in \mathcal{P}_3(\mathbb{R})$ such that for every $p \in \mathcal{P}_3(\mathbb{R})$ we have*

$$p(t) = \int_0^1 p(x)q_t(x)dx.$$

(d) *What is the polynomial $q_{1/2}$?*

Problem 5. *Consider the data set $\{(x_i, y_i)\}$ of n points displayed in Figure 0.1 obtained by observing the output y of some system given an input x. In this problem we will try to build a "model" for it, which predicts what the output y will be given an input x.*

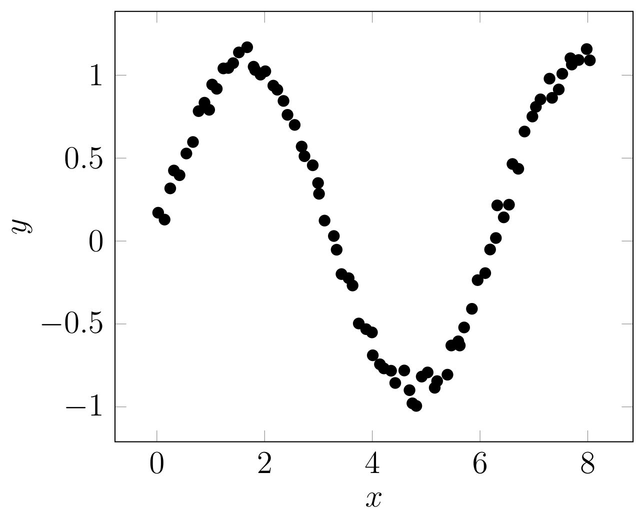

FIGURE 0.1. A data set.

One way to do so is using linear regression: we try to find the coefficients $a, b \in \mathbb{R}$ of a line $y = ax + b$ which "best fits" the data. Here the "best fit" is the one that minimizes the sum of squared errors

$$\sum_{i=1}^{n}(y_i - ax_i - b)^2. \tag{0.1}$$

To do so, one would take

$$A = \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \qquad \vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

and find

$$\begin{bmatrix} a \\ b \end{bmatrix} = (A^T A)^{-1} A^T \vec{y}.$$

(a) *Do you think this yields a good model? You do not need to do any computations or give any proofs.*

(b) *Given that the data set in Figure 0.1 looks like a sine wave, it might be appropriate to try to find the coefficients $a, b \in \mathbb{R}$ so that the graph $y = a\sin(x) + b$ minimizes the sum of squared errors*

$$\sum_{i=1}^{n}(y_i - a\sin(x_i) - b)^2.$$

*Explain with justification how to modify linear regression to do this. (*Hint: change the matrix A.*)*

One drawback of the approach of part (b) is that you have to guess the function beforehand. You will now look at the method which doesn't require this: locally estimated scatterplot smoothing.

To construct this model, we first need to discuss a variant of linear regression. In this variant, we say that the "best fit," instead of minimizing (0.1)*, minimizes the following* weighted *sum of squared errors*

$$\sum_{i=1}^{n} \mu_i (y_i - a x_i - b)^2,$$

for some weights $\mu_i \in (0, \infty)$ *for* $1 \leq i \leq n$ *picked beforehand. The result is* weighted linear regression.

(c) *Find an inner product* $\langle -, - \rangle$ *on* $\mathbb{R}^n$ *so that doing weighted linear regression amounts to finding* $\vec{c} = \left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right] \in \mathbb{R}^2$ *satisfying* $p_{\mathrm{im}(A)}(\vec{y}) = A\vec{c}$*, where* $p_{\mathrm{im}(A)}$ *is orthogonal projection with respect to* $\langle -, - \rangle$ *instead of the dot product.*

(d) *Prove that for weighted linear regression we have*

$$\begin{bmatrix} a \\ b \end{bmatrix} = (A^T D A)^{-1} A^T D \vec{y} \qquad \text{with } D = \begin{bmatrix} \mu_1 & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \mu_{n-1} & 0 \\ 0 & 0 & \cdots & 0 & \mu_n \end{bmatrix}.$$

Locally estimated scatterplot smoothing is then given by the following procedure. Given an input $x \in \mathbb{R}$*, we do weighted linear regression with weights* $\mu_i(x) = (1 - c|x_i - x|^3)^3$ *to find coefficients* a, b *(so these depend on* x*) and say our prediction is for the output is* $ax + b$*. Here* c *is a constant to be chosen.*

(e) *Explain why the inventors of locally estimated scatterplot smoothing might have picked the weights to be* $\mu_i(x) = (1 - c|x_i - x|^3)^3$*, and give guidelines for choosing* c*. You do not need to do any computations or give any proofs.*