Some Review Exercises on Units 1-5

Many taken/adapted from [VMLS] or [LALFD]

(1) Consider a function $f : \mathbb{R}^n \to \mathbb{R}^m$ that is linear. Show that there exists a unique $m \times n$ matrix A such that $f(x) = Ax$.

(2) Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ with $f(x) = x_1 x_2 + e^{x_1^2}$. Find the gradient of f and obtain the linear approximation at $x = (1, 1)$.

(3) Verify that for any affine function $f : \mathbb{R}^n \to \mathbb{R}$ with $f(x) = a^T x + b$:

$$f(x) = f(0) + \sum_{i=1}^{n} x_i \left(f(e_i) - f(0)\right).$$

Generalize to an affine function $f : \mathbb{R}^n \to \mathbb{R}^m$.

(4) Show: $||x + y|| = \sqrt{||x||^2 + 2x^T y + ||y||^2}$.

(5) Remember that

$$\mathrm{rms}(x) = \frac{||x||}{\sqrt{n}} \qquad \mathrm{avg}(x) = \frac{\mathbf{1}^T x}{n} \qquad \mathrm{std}(x) = \frac{||x - n^{-1}\mathbf{1}^T x||}{\sqrt{n}}.$$

Now show that,

$$\mathrm{std}(x)^2 = \mathrm{rms}(x)^2 - \mathrm{avg}(x)^2.$$

(6) Prove the Cauchy-Schwarz inequality: $|a^T b| \leq ||a||\ ||b||$.

(7) Let a and b be two different n-vectors. The line passing through a and b is given by the set set of vectors of the form $(1 - \theta)a + \theta b$ where θ is a scalar that determines the particular point on the line. Now let x be any n-vector. Find a formula for the point p on the line that is closest to x. This point is called the projection of x onto the line. Show that $(p - x) \perp (a - b)$.

(8) Let $x_1, \ldots, x_L$ be a collection of n-vectors. Consider,

$$J(z) = \sum_{i=1}^{L} ||x_1 - z||^2.$$

Find the vector z that minimizes $J(z)$ and prove your result. Describe how this relates to the k-means algorithm.

(9) Prove that an orthonormal (or even orthogonal) set of vectors is linearly indepndent and say what this says about the number of vectors in the set.

(10) Say you have two different basia (plural of basis) for $\mathbb{R}^n$. The basis $\mathcal{B}_1$ is orthonomral and the basis $\mathcal{B}_2$ is not. You have a vector x and want to expand x in two different ways. Once in terms of $\mathcal{B}_1$ and once in terms of $\mathcal{B}_2$. How would you go about finding such expansions in an efficient manner. Which is easier/quicker?

(11) Carry out the Gram-Schmidt algorithm on the vectors $a_1 = (1,2,3)$, $a_2 = (1,0,3)$, $a_3 = (3,2,6)$ and $a_4 = (1,1,1)$.

(12) Repeat now without a_4 and with $a_3 = (2,2,6)$ instead. Suggest how to modify the algorithm so that in cases such as the first, it won't terminate but rather flag the vectors a_3 and a_4 to be linear combinations of the previous vectors.

(13) Say we run the Gram-Schmidt algorithm twice. First on a given independent set of vectors $a_1, \ldots, a_k$ which gives the vectors $q_1, \ldots, q_k$. Then we run again on $q_1, \ldots, q_k$ as input to obtain $z_1, \ldots, z_k$. What can you say about $z_1, \ldots, z_k$?

(14) Take A as an $m \times n$ matrix. Say you want to exchange rows i and j of A by multiplying by a matrix P. What is P and would you left-multiply or right-multiply? What if you wanted to exchange collumns?

(15) Let A be an $m \times n$ matrix. Consider the stacked matrix S defined by,

$$S = \begin{bmatrix} A \\ I \end{bmatrix}.$$

When does S have lineary independent collumns? When does S have linearly indepdnent rows? Your answer can depend on m, n or whether or not A has linearly indpendent columns or rows.

(16) Let $f : \mathbb{R}^p \to \mathbb{R}^m$ and $g : \mathbb{R}^n \to \mathbb{R}^p$ and define $h(x) = f\big(g(x)\big)$. The is $h : \mathbb{R}^n \to \mathbb{R}^m$ and,

$$h(x) = f(g_1(x), \ldots, g_p(x)).$$

From the chain rule, we have that,

$$\frac{dh_i}{dx_j}(z) = \sum_{\ell=1}^{p} \frac{df_i}{dy_\ell}\big(g(z)\big)\frac{dg_\ell}{dx_j}(z). \qquad \text{for } i = 1, \ldots, m, \;\; j = 1, \ldots, n.$$

Show that this can be described concisely via Jacobian matrices as follows:

$$Dh(z) = Df\big(g(z)\big)Dg(z).$$

Now show that the first or Taylor approximation of h at z can be written as ,

$$\hat{h}(x) = f\big(g(z)\big) + Df\big(g(z)\big)Dg(z)\,(x - z)$$

(17) Suppose the columns of A are orthonormal. Show that $||Ax|| = ||x||$.

(18) Let $a_1, \ldots, a_n$ be the collumns of the $m \times n$ matrix A. Suppose the collumns all have norm 1 and for $i \neq j$ the angle between a_i and a_j is 60 degrees. What can be said about the gram matrix $G = A^T A$.

(19) A student says that for any square matrix A,

$$(A + I)^3 = A^3 + 3A^2 + 3A + I.$$

Is she right? Explain why? Can you generalize this claim?

(20) Let U and V be two orthogonal $n \times n$ matrices. Show that the matrix UV is orthogonal. Show also that the $2n \times 2n$ matrix,

$$W = \frac{1}{\sqrt{2}} \begin{bmatrix} U & U \\ V & -V \end{bmatrix}$$

is orthogonal.