Midterm 1

- Don't panic.
- You *may* use your computer and/or calculator.
- You *may* use your notes, book, another person, or any other content resource.
- There is a very high level of professionalism expected with this assessment; make sure the final deliverable you submit is clear, concise and properly documented.
- This assessment is due **Monday, 11 March at 5PM** outside my office, Babson 316.
- Make a study copy of this assessment before you submit it.

(1) Sometimes a single independent variable isn't enough to create a dependable model of a given system. For an example, imagine that you run a small ice cream shop on the coast of Maine. There are two main drivers for your sales: daily temperature and median customer income. Suppose we have a model with two independent variables u, representing the average daily temperature in July in your town, and v, representing the median income of customers who purchased from you in July. You've been collecting data over several years. The results can be seen in Table 1

Year	**Total Sales**	**Average Temp.**	**Median Income**
2009	27.93	86.92	30.11
2010	28.29	88.51	31.48
2011	29.70	88.01	32.03
2012	31.09	87.05	33.34

Table 1: Data collected for the total sales (thousands of dollars), average temperate (degrees Fahrenheit), and median household income (thousands of dollars) for July of the indicated year

(a) Is there a unique multilinear model $s(u, v) = \beta_0 + \beta_1 u + \beta_2 v$ that perfectly fits your data?

(b) We could also allow for the variables u and v to interact multiplicatively through the model $s(u, v) = \beta_0 + \beta_1 u + \beta_2 v + \beta_3 uv$. Is there a unique model of this form that perfectly fits your data?

(c) An even more general model might be $s(u, v) = \beta_0 + \beta_1 u + \beta_2 v + \beta_3 uv + \beta_4 u^2 + \beta_5 v^2$. Is there a unique model of this form that perfectly fits your data?

(2) One of the tools used in data mining is logistic regression, which takes a collection of observations about certain probabilities and attempts to construct the underlying cumulative density function. The logistic function in this case is

$$\pi(x) = \frac{e^{\beta_0+\beta_1 x}}{e^{\beta_0+\beta_1 x}+1},$$

where β_0 and β_1 are the parameters to be estimated. Notice that $\pi(x)$ is between 0 and 1 for every x. Moreover, we see that $\pi(x) \to 1$ as $x \to \infty$ and $\pi(x) \to 0$ as $x \to -\infty$. So $\pi(x)$ seems like a pretty good candidate for a CDF. Then $\pi(x)$ is the probability that some random variable X has value less than or equal to x.

(a) For an example, let's turn our attention to income distribution in the United States. Census data indicate that 9.54% of American households earn less than or equal to $10,000 per year, and 87.71% of American households earn less than or equal to $100,000 per year. Use these data to compute an approximate cumulative density function for the distribution of American household income. Hint: consider $\pi(x)/(1 - \pi(x))$.

(b) According to your model, what is the percentage of American households earning less than $200,000 per year?

(c) How much does a household have to earn in order to be in the top 1%?

---

(3) Let $f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3$ be an arbitrary degree 3 polynomial. If you think about it, the coefficients completely determine this polynomial, and so we can package the all the information about $f(x)$ in a convenient vector form $\mathbf{f} = [a_0, a_1, a_2, a_3]^T$. Why is this convenient, you ask? Well, we can represent common operations like differentiation with a simple matrix, as you'll have the chance to show below.

(a) Construct a 4×4 matrix $\mathbf{D}$ which "differentiates" our vector $\mathbf{f}$, that is, a matrix $\mathbf{D}$ such that $\mathbf{Df} = \mathbf{f}'$, where $\mathbf{f}'$ is the vector containing the coefficients of $f'(x)$.

(b) Construct another 4×4 matrix $\mathbf{E}$ which takes the second derivative of our vector $\mathbf{f}$.

(c) Show that $\mathbf{D}^2 = \mathbf{E}$. Explain as clearly and concisely as possible why these equalities hold.

(4) After a lot of data collection, you've assembled the matrix

$$\Delta P = \begin{bmatrix} 2.90 & 1.84 & 0.56 & 1.41 \\ 1.72 & 2.54 & 0.70 & 1.23 \\ 1.66 & 1.92 & 1.69 & 1.39 \\ 1.72 & 1.76 & 0.62 & 2.05 \end{bmatrix},$$

where $\Delta P(r, c)$ represents the change in price of product r realized after a \$1 increase in the value added of product c. In the parts that follow, use Leontief models.

(a) How many more units of product 2 are required if the final demand for product 3 increases by 1 unit?

(b) How many units of product 3 are required to make 1 unit of product 4?

(c) What is the total cost to make 1 unit of product 1 in terms of the price per unit set by each industry?

(d) The gross domestic product of an economy is the value of all goods produced. In a Leontief model, we can express the GDP in two ways. Show that $\mathbf{p}^T\mathbf{d} = \mathbf{v}^T\mathbf{x}$.