Problem 1. Matlab: defining a function and using subplots (20)

Define a function

\[ h(x; w) = \frac{1}{1 + e^{-wx}} \]

For each of the four values of \( w = \{0, .015, .1, 1000\} \) plot \( h(x; w) \) over \( x = [-70 : 70] \). Using the `subplot` command, place these plots in a 2-by-2 grid. Title each subplot with the value of \( w \) used in that subplot. Standardize the axes across subplots with the command `axis` for ease of comparison.

Print out this figure. Notice from the plots that a single function can behave like a constant, a line, a sigmoid, or a step function, depending on the value of the parameter \( w \). This family of functions is known as logistic functions. You will see them again frequently later in the course (not the least in logistic regression!).

(Solution)
Problem 2. Matlab: polynomial curve-fitting (30)

(a) (15) You will generate data from a quadratic function of the form

\[ y = ax^2 + bx + c \]

First generate 10 values for \( x \) between -10 and 10 using the command \( x = \text{rand}(10,1)*20-10 \). Use the command \text{sort} to sort \( x \) into ascending values. Then generate the corresponding \( y \)-values when \( a = 1, b = -4, c = -21 \). Now produce a “noisy” version of the \( y \) values, using the \text{normrnd} command to generate Gaussian noise with \( \mu = 0, \sigma = 10 \), which you add to \( y \), yielding the noisy \( yn \).

Use the command \text{plot} to plot \( y \) versus \( x \) with a blue line. Use the command \text{hold} and \text{plot} again to plot \( yn \) versus \( x \) with green circles.

Use the command \text{regress} to find the best 0th-, 1st-, 2nd-, 5th-, and 10th-order polynomial fits to \( yn \) as a function of \( x \). Use \text{plot} to display the polynomial fits with different-colored
dashed lines, alongside $y$ and $yn$. Use `legend` to label $y$, $yn$, and the various polynomial fits. Print out this figure.

(Solution)

(b) (15) Compute the root-mean-square error (defined in lecture 1) for each polynomial fit. In a new figure, plot this error (known as “training error”) against the order of the polynomial. Then generate 10 more values for $x$, using the command $x2 = \text{rand}(10,1)*20-10$. Compute the “true” $y2$ values using the equation in (a). Now compute the root-mean-square error between the polynomial fit obtained from (a) and $y2$. Use `hold` and `plot` to plot this “test error” as a function of the order of the polynomial. How does it compare to the training error’s dependence on the order of the polynomial? Print out this figure.

(Solution)
Problem 3. Math review. Prove that the following are true (20)

(a) (5) Cyclic permutations are allowed inside the trace:

\[ \text{Tr}(BCD) = \text{Tr}(DBC) = \text{Tr}(CDB) \]

(Solution)
\[ \text{Tr}(BCD) = \sum_i \sum_j \sum_k B_{ij} C_{jk} D_{ki} \]
\[ = \sum_k \sum_i \sum_j D_{ki} B_{ij} C_{jk} \]
\[ = \text{Tr}(DBC) \]
\[ = \sum_j \sum_k \sum_i C_{jk} D_{ki} B_{ij} \]
\[ = \text{Tr}(DCB) \]

(b) (5) Recall trace of a scalar is a scalar, show the following:
\[ \text{Tr}(xx^\top B) = \text{Tr}(x^\top Bx) = x^\top Bx \]

(Solution)
\[ x^\top Bx = (x_1, ..., x_k) B_{k\times k} (x_1, ..., x_k)^\top \]
\[ = \text{Tr}(x^\top Bx) \]
\[ = \sum_i \sum_j \sum_k x_{ij} B_{jk} x_{ki} \]
\[ = \sum_k \sum_i \sum_j x_{ki} x_{ij} B_{jk} \]
\[ = \text{Tr}(xx^\top B) \]

(c) (10) Some calculus review
\[ \frac{d}{dz} \left[ \frac{1}{(b + cz)} \exp(az) + e \right] = \exp(az) \left( \frac{a}{b + cz} - \frac{c}{(b + cz)^2} \right) \]

Note that \( a, b, c \) and \( e \) are constants and \( 1/u = u^{-1} \). Hint: recall the chain rule.

(Solution)
\[ \frac{d}{dz} \left[ \frac{1}{(b + cz)} \exp(az) + e \right] = \frac{d}{dz} \left[ \frac{1}{(b + cz)} \exp(az) \right] \]
\[ = \exp(az) \frac{d}{dz} \frac{1}{(b + cz)} - \exp(az) \frac{d}{dz} \frac{1}{(b + cz)^2} \]
\[ = a \exp(az)(b + cz) - \exp(az)c \]
\[ = \exp(az) \left( \frac{a}{b + cz} - \frac{c}{(b + cz)^2} \right) \]
Problem 4. Basic probability theory (30)

(a) (5) Show that closure under complementation and closure under countable union together imply closure under countable intersection:

\[ E_1, E_2, \ldots \in \mathcal{F} \Rightarrow \cap_i E_i \in \mathcal{F} \]

Hint: use the set theory version of DeMorgan’s Law (http://en.wikipedia.org/wiki/De_Morgan’s_laws)

(Solution)

\[ \cap_i E_i = (\cup_i (E_i)^c)^c \]

Since \( E_i \in \mathcal{F} \) then \( (E_i)^c \in \mathcal{F} \) thus \( \cup_i (E_i)^c \in \mathcal{F} \)

therefore \( (\cup_i (E_i)^c)^c \in \mathcal{F} = \cap_i E_i \in \mathcal{F} \)

(b) (5) Suppose we are casting a die, and we know we can assign a probability to each atomic event (e.g. \( P(x=i) = 1/6 \) \( \forall i \in \{1, 2, 3, 4, 6\} \)). From that we can compute the probability of many other events, like the event of the die coming up odd, or being greater than 5. Including the null event = {} and the all-inclusive event of the sample space itself \( \Omega = \{1, 2, 3, 4, 5, 6\} \), how many total measurable events are there?

(Solution)

Definition: A pair \((S, \mathcal{F})\) is called a measurable space, and any subset in \( \mathcal{F} \) is called a measurable event.

For every elementary event in sample space \( S = \Omega = \{1, 2, 3, 4, 5, 6\} \) there exists a complementary event such that event \( A = \{1\}, A^c = \{2, 3, 4, 5, 6\} \).

\[ \mathcal{F} = \{\emptyset, \{1\}, \{2, 3, 4, 5, 6\}, \{2\}, \{1, 3, 4, 5, 6\}, \ldots, \{1, 2, 3, 4, 5, 6\}\} \]

Hence, there are \( 2^6 = 64 \) measurable events.

(c) (5) Monotonicity: for any two events \( A_1, A_2 \in \mathcal{F} \), show that \( A_1 \subseteq A_2 \Rightarrow P(A_1) \leq P(A_2) \).

(Solution)

Suppose \( A_1 \) is a subset of \( A_2 \) then

\[ A_2 = A_1 \cup (\bar{A}_1 \cap A_2) \]

mutually exclusive \( A_1 \cap (\bar{A}_1 \cap A_2) = \emptyset \)

\[ P(A_2) = P(A_1) + P(\bar{A}_1 \cap A_2) \]

\[ 0 \leq P(\bar{A}_1 \cap A_2) \leq 1 \]

so \( P(A_1) \leq P(A_2) \)

(d) (5) Sub-additivity: For any events \( A, A_i \in \mathcal{F} \), show that \( A \subseteq \cup_i A_i \Rightarrow P(A) \leq \sum_i P(A_i) \).

(Solution)

\[ P(A) \leq P(\cup_i A_i) \]
\[ P(\cup_i A_i) \leq \sum_i P(A_i) \]

\[ B_i = A_i \cap (\cup_{j=1}^{i-1} A_j)^c \subseteq A_i \]

then \( B_1, B_2, \ldots \) are disjoint events.

\[ \cup_i B_i = \cup_i A_i, \quad P(B_i) \leq P(A_i) \]

\[ P(\cup_i A_i) = P(\cup_i B_i) = \sum_i P(B_i) \leq \sum_i P(A_i) \]

\[ P(A) \leq P(\cup_i A_i) \leq \sum_i P(A_i) \]

\[ P(A) \leq \sum_i P(A_i) \]

(e) (5) Bayes’ Theorem:

\[ P(B|A) = \frac{P(A|B)P(B)}{\sum_b P(A|B=b)P(B=b)} \]

(Solution)

\[ P(A \cup B) = P(A|B)P(B) \]

\[ = P(B|A)P(A) \]

\[ P(B|A) = \frac{P(A \cup B)}{P(A)} \]

\[ = \frac{P(A|B)P(B)}{\sum_b P(A|B=b)P(B=b)} \]

(f) (5) Show that if the events \( A \) and \( B \) are independent, then \( P(A|B) = P(A) \) and \( P(B|A) = P(B) \).

(Solution)

\[ P(A \cup B) = P(A)P(B) \]

\[ P(A|B) = \frac{P(A \cup B)}{P(B)} \]

\[ = \frac{P(A)P(B)}{P(B)} \]

\[ = P(A) \]

\[ P(B|A) = \frac{P(B \cup A)}{P(A)} \]

\[ = \frac{P(B)P(A)}{P(A)} \]

\[ = P(B) \]