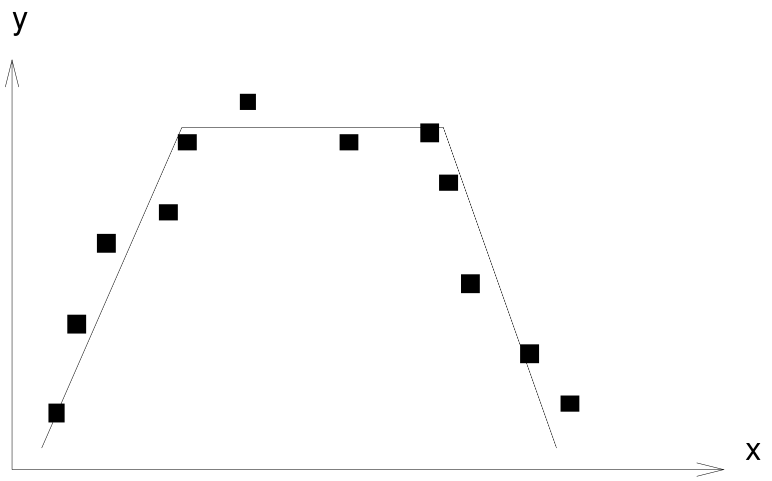
# 6.824 Artificial Intelligence, Fall 1992
## Problem Set 3

Due: Beginning of class Wednesday, October 13

**Problem 1.** This problem is on MDL. Suppose that we want to fit a curve to a set of points. Each point is represented by a pair $<x,\ y>$. Suppose our "curves" are actually piecewise linear functions such as the one shown in following diagram.

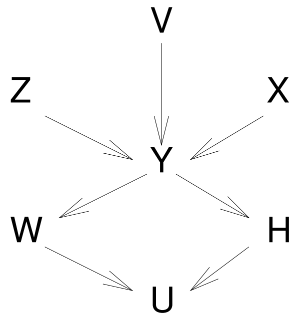



In general we would like to if a piecewise linear function to data (the function shown in the diagram is a piecewise linear function with three linear regions). Let $f$ be a piecewise linear function and let $D$ be a set of points $<x_1,\ y_1> \ldots <x_n,\ y_n>$. Assume that the data is generated by first selecting the function $f$ and then generating each $y_i$ by adding Gaussian noise to $f(x_i)$.

**part a.** Use the MDL principle to define a joint probability distribution over piecewise linear functions and data. You should allow for piecewise linear functions of an arbitrary number of linear regions. Give an expression for $-log_2(P_r(f \wedge D))$ under the distribution you define (your expression may contain an arbitrary additive constant).

**part b.** Define a graph whose nodes are the data points $<x_i,\ y_i>$ and such that there is an arc from $<x_i,\ y_i>$ to $<x_j,\ y_j>$ whenever $x_i < x_j$. Assign weights to the arcs in such a way that finding a piecewise linear function $f$ that maximizes $P_r(f|D)$ under the distribution given in part a corresponds to finding a minimum weight path from the data point with smallest $x$ value to the data point with largest $x$ value. Based on this graph show that one can find the optimal piecewise linear function in $O(n^2 \log n)$ time where $n$ is the number of data points. You may use the fact that finding the shortest path between two nodes in a weighted directed graph with positive weights and with $n$ nodes and $e$ edges can be done in $O(e \log n)$ time.

**Problem 2.** Consider the following Bayesian network.




Let $I(X,\, Z,\, Y)$ mean that variables $X$ and $Y$ are independent given a value for $Z$, i.e., $P_r(X = x_i \wedge Y = y_k | Z = z_j) = P_r(X = x_i | Z = z_j) P_r(Y = y_k | Z = z_j)$. (It is important that in this problem $X$, $Z$ and $Y$ are single variables rather than sets of variables as is done by some authors.) List all of the independence statements of this form implied by the structure of the above network, i.e., that are true for all Bayesian networks with this topology. Explain your answer.

**Problem 3.** Consider the Bayesian network consisting of the following linear sequence of nodes.

$P_1 \rightarrow P_2 \rightarrow \ldots \rightarrow P_T$

Suppose that each node is a Boolean variable and that for arrow $P_t \rightarrow P_{t+1}$ the matrix $M^{P_{t+1}}$ is such that there is a $\frac{3}{4}$ chance that $P_{t+1}$ has the same value as $P_t$. More precisely, for each node $P$ other than $P_1$ we have $M^P_{T|T} = M^P_{F|F} = \frac{3}{4}$ and $M^P_{T|F} = M^P_{F|T} = \frac{1}{4}$. Also suppose that the prior probability of $P_1$ is $\frac{1}{2}$. Let $O$ be the observation that $P_1 = T$ and $P_T = T$.

**Part a.** For the observations $O$ described above give a recurrence relation for computing $\beta_{P_t}(T)$ and $\beta_{P_t}F$ from $\beta_{P_{t+1}}(T)$ and $\beta_{P_{t+1}}F$

**Part b.** Solve the recurrence you got in part a to get a closed form solution for $\beta_{P_t}(T)$ and $\beta_{P_t}(F)$ as a function of $t$.

**Part c.** Give a recurrence relation for computing $\alpha_{P_{t+1}}(T)$ from $\alpha_{P_t}(T)$. Solve this recurrence to get a closed form solution for $\alpha_{P_t}(T)$ as a function of $t$.

**Part d.** Combine parts b and c to give a closed form formula for $P_r(P_t|O)$ as a function of $t$.

**Problem 4.** Consider a Bayesian network which is an $n$ level binary tree with arcs directed from parents to children and ending in the $2^n$ leaves of the tree. Each node in this tree is a Boolean variable where the probability that a node has the same value as its parent is $\frac{3}{4}$. The probability that the root node is true is $\frac{1}{2}$. Suppose that we observe that all the leaf nodes have value $T$. We use $P_n$ to represent a node that is $n$ arcs above the leaf nodes. For example $P_0$ is a leaf node and $P_2$ is a node two levels above the leaf nodes

**part a.** Give a recurrence relation for computing $\beta_{P_{n+1}}(T)$ and $\beta_{P_{n+1}}(F)$ from $\beta_{P_n}(T)$ and $\beta_{P_n}(F)$.

**part b.** Using your answer from a, compute $\beta_{P_n}(T)$ and $\beta_{P_n}(F)$ for values of n from 0 to 4.

**part c.** Give a closed form solution for $\alpha_{P_n}(T)$ as a function of $n$.

**part d.** Using your results from b and c, and some additional informal arguments, give a value for the limit of $P_r(P_n|O)$ as $n$ goes to infinity, i.e., for arbitrarily deep trees all of whose leaves are observed to be true.