Assignment no-2.

1. Describe the need for introdⁿ of non-linear funⁿ in regression of pattern recognition using relevant representaⁿ.

→ Linear models must follow one very particular form.

Dependent variable = constant + parameters because all terms are either the constant or a parameter multiplied by an independent variable (IX). A linear regression equations simply sums the terms. While the model must be linear in the parameters you can raise an independent variable by an exponent to fit a curve.

Being a linear funⁿ of input variable x, limits the usefulness of the funⁿ.

This is because ~~of~~ most of the observaly that may be, to be a linear combinaⁿ of fixed non-linear funⁿ of the input variable.

Why use non-linear regression?

1. Transformation is necessary to obtain variance homogeneity, but transformation destroy linearity.

PAGE NO.:
DATE: / /

2. Linearity does not fit, and the transformation seems to destroy other parts of the model assumptions, e.g. the assumpⁿ of variance homogeneity.

3. Theoretical knowledge (e.g from kinetics or physiology) indicates that the proper relation is intrinsically non-linear.

4. Interest is in functions of the parameters that do not enter linearly in the model (ex- kinetic rate constants or ED50 in dose-response studies).

Expression for non-linear regression-

If we assume that the non-linear function of input variable is $\varphi(x)$, then we can re-write the original function as-

$$y(x,w) = w_0 + w_1\varphi(x_1) + w_2\varphi(x_2) + \dots + w_D\varphi(x_D)$$

summing it up, we will have.

$$y(x,w) = w_0 + \sum_{j=1}^{M-1} w_j \phi(x)$$

where φ are known as basis functions.

PAGE NO.:
DATE: / /

Example-

Quantification of the Reticuloendothelial cell system (RES) of the liver.

concentration measurements y_i, over the liver, following a bolus injection of radioactive tracer.

~~first~~ order kinetics implies

$$c(t) = \beta(1 - e^{-4t})$$

~~No~~ transform to linearity possible

$$y_i = \beta(1 - e^{-4t_i}) + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2)$$

Thus this cannot be solved by linear model hence we go for non-linear function regression in such problems.

Analyze the Bayesian classification process using real time example in detail?

Bayesian classification ~~process using real~~ is based on Bayes theorem.
Bayesian classifiers are the statistical classifiers.
Baysian classifiers can predict class membership probabilities such as the probability that a given tuple belongs to a particular class.

Baye's Theorem —

Baye's theorem is named after thomas Bayes. There are two types of probabilities :—

Posterior Probability [$P(H|X)$]
Prior probability [$(P|H)$]

where x is data tuple and H is some hypothesis.
According to Baye's theorem:

$$p(H|x) = p(x/H)\,p(H) \,/\, p(x)$$

likelihood   class prior probability

$$p(c|x) = \frac{P(x|c)\,P(c)}{P(x)}$$

posterior probability   predictor prior probability.

PAGE NO.:
DATE: / /

Bayesian Belief n/w

Bayesian Belief n/w specify joint conditional probability distributions. They are also known as belief n/ws, Bayesian n/ws or probabilistic n/w.

- A Belief n/w allows class conditional independancies to be defined betn subset of variables.

It provides a graphical model of casual relationship on which learning can be performed.

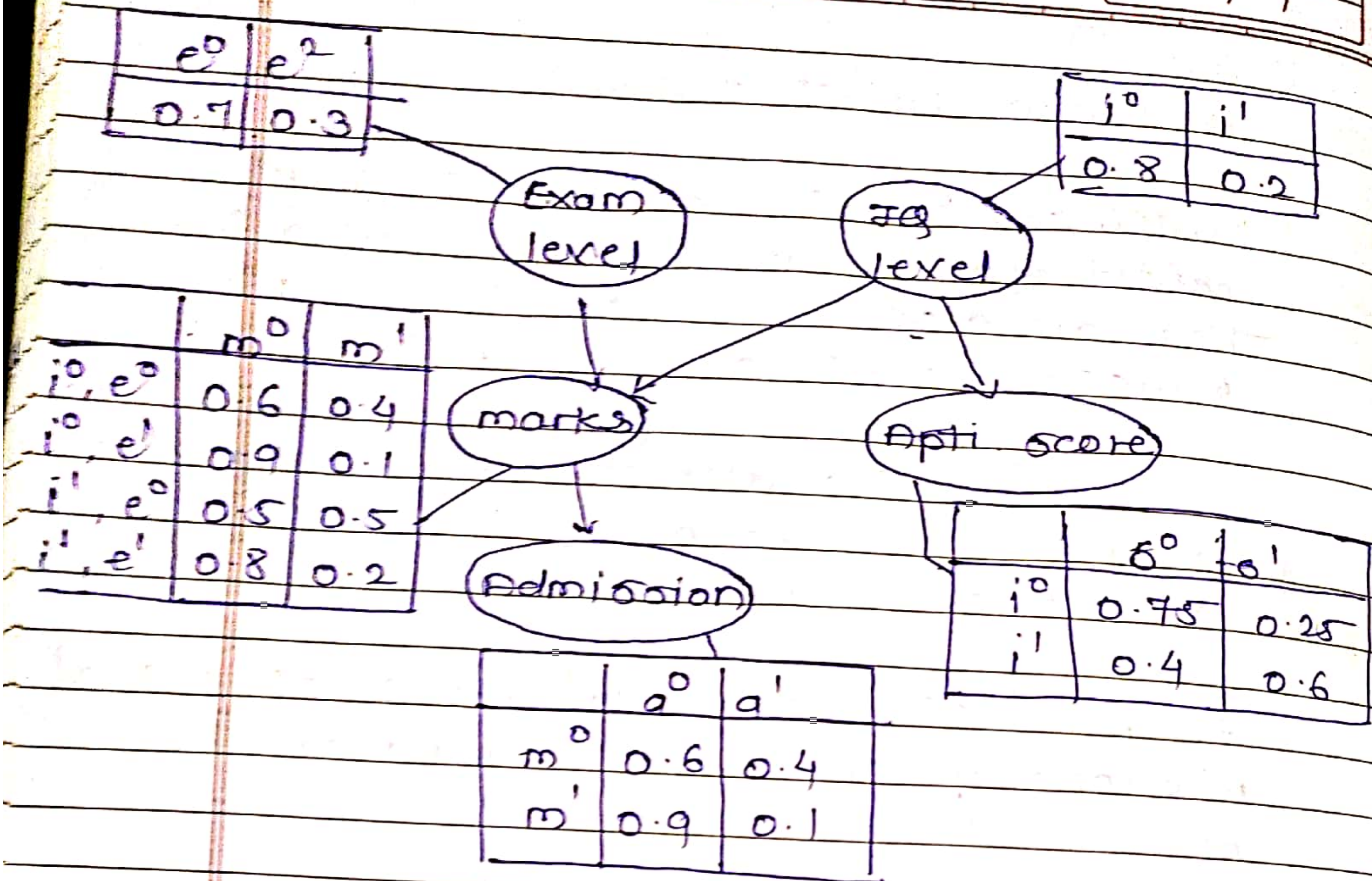
- We can use a trained Bayesian n/w for classification.

There are components that define a Bayesian belief n/w —

- Directed acyclic graph - The distribn is represented by nodes in graph, which are variables and edge betn the nodes describe the conditional dependancies.

Requires that graph is acyclic (no directed cycles)

2 components to a Bayesian n/w.



e^0	e^1
0.7	0.3

i^0	i^1
0.8	0.2

	m^0	m^1
i^0, e^0	0.6	0.4
i^0, e^1	0.9	0.1
i^1, e^0	0.5	0.5
i^1, e^1	0.8	0.2

	s^0	s^1
i^0	0.75	0.25
i^1	0.4	0.6

	a^0	a^1
m^0	0.6	0.4
m^1	0.9	0.1

Above I've represented this distribⁿ through a DAG and a conditional probability table. We can now calculate the joint probability distriⁿ of these ~~v~~ variable ie, the product of conditional probabilities.

$$p(a, m, i, e, s) = p(a|m)\, p(m|i, e)\, p(i)\, p(e)\, p(s|i)$$

here,

- $p(a|m)$ represents the conditional probability of a student getting an admission based on his marks.

- p(m|I, e) represents the conditional probability of the student's marks, given his ~~I/Q level~~ IQ level and exam level.

- p(i) denotes the probability of his IQ level (high or low)

- p(e) denotes the probability of the exam level (difficult or easy)

- p(s|i) denotes the ~~probability~~ conditional probability of his aptitude scores, given his IQ level.

The DAG clearly shows how each variable (nodes) depends on its parent node, i.e, the marks of the student depends on the exam level (parent node) and IQ level (parent node) similarly, the aptitude score depends on the exam level (parent node) and finally, his admission into a university depends on his marks (parent node). This relationship is represented by the edge of the DAG.

If you notice carefully, we can see a pattern here. The probability of a random variable depends on his parents. Therefore, we can formulate Bayesian ~~this~~ as:

$$p(x_1, \ldots, x_n) = \prod^{n}_{i=1} p(x_i \mid parents(x_i))$$

Where, x_i denotes a random variable whose probability depends on the probability of the parent node (-).