Super AGI Assignment: SANGAM KUMAR
2022AIB2671

Q1
When you duplicate n feature into n+1 and retrain a model then the weights of the logistic regression model are likely to be distributed in similar manner b/w $W_n \leftrightarrow W_{n+1}$ since they both represent duplicated information.

Weight are influenced by ① data
② training Algorithm

So the duplicated features may share the impact on the model.

But we can say the relationship b/w W_{new} & $W_{new\,n+1}$ is likely to involve similarity in magnitudes, but the exact nature of this relationship depends on the data and training process.

Q2
E is better than A with over 95% confidence, B is worse than A with over 95% confidence. You need to run the test for longer to tell where C & D compare to A with 95% confidence.

Ans 3 Given m training examples & n features
feature vectors are sparse,

Objective: To find approximate Computational Cost of each gradient descent iteration.

predict

Ans 4 The accuracy of classifier V_2 will ~~keep~~ likely vary dependending on the method used to generate the addition of training data:

① ~~Poo~~ It focus on instances where V_1 classifier's decision is uncertain or close to the decision boundary, potentially helping V_2 learn more different distinctions

② Involves random selection of labeled stories, provides a diverse set of examples of training.

③ targets stories where V_1 classifier makes mistakes and select those that are farthest away from the decision boundary, possibly capturing cases where the model is confidently wrong.

Each approach has its advantages and considerations.

① + ② → address case where V_2 will perform better than V_1.

Q5 (i) Given distribution is a binomial distribution

n times toss $\longrightarrow$ k times head with each time probability P

(a) Maximum likelihood estimate (MLE)

as given $\rightarrow$ n is given but want to estimate P

K : binomial parameters n (known) + P (unknown)

$$L_P(k; P) = \binom{n}{k} P^k (1-P)^{n-k}$$

log of both sides

$$L_P = \log \begin{bmatrix} n \\ k \end{bmatrix} + k \log \theta + (n-k) \log (1-\theta)$$

MLE

To maximize MLE $\frac{\partial}{\partial P}(L_P) = 0$

$$0 = 0 + \frac{K}{P} - \frac{n-k}{1-P} = 0$$

$$K(1-P) = P(n-k)$$

$$K - \cancel{KP} = Pn - \cancel{PK}$$

$$K = Pn$$

$$\theta_{MLE} = \frac{k}{n}$$

(b) Bayesian Estimate:

given that P is the probability of an event

Prior distribution is assumed to be a continous

Now formulation of problem under Binomial distribution

$$f(k/p) = \binom{n}{k} p^{k} (1-p)^{n-k}.$$

Possible prior on p. $\sim U(0,1)$.

$$P(k) = 1$$

Posterior distribution let $(p = \theta)$
small $p \rightarrow$ Probability

$$P(\theta/k) \propto \binom{n}{k} \theta^{k} (1-\theta)^{n-k}$$

$$\propto (\theta)^{k+1-1} (1-\theta)^{n-k+1-1}$$

defines a β density $(k+1, n-k+1)$

Bayes estimate of $\theta = \dfrac{k+1}{n-k+1}$

finally answer

$$P = \frac{k+1}{n-k+1}.$$

(c) find Maximum a posteriori (MAP)

Assume that prior is uniform

$P(k) = 1$ Prior on $P \sim U(0,1)$

MAP estimates corresponds to the mode of the posterior distribution which is the value of P that maximizes $\log(P(P|X))$

$$MAP \propto MLE * PRIOR$$

$$PRIOR - P(K) = 1 \text{ for } [0,1].$$

$$MAP \propto MLE$$

$$MAP = K/n$$