Weekly report of lessons

Name: Mayank Kumar
Roll No: 19CS30029
The week: 06-09-2021 to 12-09-2021

The topics covered:

Evaluating hypothesis: Evaluating accuracy of a hypothesis, Expectation and Variance, Error of hypothesis, Bernoulli Distribution, Probabilistic Analysis, Probable range of estimate, Confidence Interval, Central Limit Theorem, k-fold Cross-Validation, Comparing Hypotheses;
Bayesian decision theory and learning: Random Process, Bayesian Interface, Maximum a Posteriori (MAP) Hypothesis, Maximum Likelihood (ML) Hypothesis, Features of Bayesian Learning, Concept Learning under Bayesian framework, Least Mean Square Error estimate as ML hypothesis

Summary topic wise:

Evaluating hypothesis

Evaluating accuracy of a hypothesis: Straightforward when data is plentiful. Two difficulties arise for limited data - Bias and Variance in the estimate

Expectation and Variance: For random variable X, probability Pr and probability distribution function p_θ
Discrete: $E[X] = \Sigma_x\, x\, Pr(X=x)$
Continuous: $E[X] = {}_{-\infty}\!\int^{\infty} x\, p_\theta\, (x)\, dx$
Variance: $Var(X) = E((X - E[X])^2) = E[X^2] - E[X]^2$

Error of hypothesis: for an instance x of distribution D, data sample S of size n, hypothesis h: X -> {0, 1}, error function: e(x,y) = 1 if x~=y otherwise 0
Sample Error, $E_S(h) = (1/n)\, \Sigma_{x \in S}\, e(f(x), h(x))$ and True Error, $E_D(h) = Pr_{x \in D}\, \{f(x) \neq h(x)\}$

Bernoulli Distribution: For random variable X with $Pr(X=1) = p$, we will have
$Pr(X=0) = 1 - p$; $E[X] = p$; $Var(X) = p\ (1-p)$

Probabilistic Analysis: Probability of error for a sample: $E_D(h) = p$
Probability of r errors in n samples: $({}^nC_r)\ p^r\ (1-p)^{n-r}$
$E(r) = np$ and $var(r) = np(1-p)$
$E(r/n) = p$ and $var(p) \simeq var(r/n) = p(1-p)/n$
Inductive bias: a set of assertions
Bias of estimate: a numerical quantity = E(estimate) - true-estimate-value

Probable range of estimate: Given r errors in n ($\geq$ 30) samples, $E_s(h) = r/n$
Given no other information, $E_D(h) = E_s(h)$
For approx 95% prob,. E_D(h) lies in confidence interval -> $E_S(h) \pm 1.96\ \sqrt{(E_S(h)\ (1-E_S(h))\ /\ n)}$

Confidence Interval: Interval containing true value with probability 95% is N% confidence interval. For a normal distribution N(μ, σ), N% C.I. = $\mu \pm Z_n\sigma$

Central Limit Theorem: For a set of random independent variables $Y_1, Y_2, \ldots, Y_n$, Mean = Y'.
As n -> ∞, distribution governing Y' approaches a Normal Distribution N(μ, σ).

k-fold Cross-Validation: Randomly divide the set of observations into k groups of approximately equal size, observe Y_i and compute the N% confidence interval.

Comparing Hypotheses:
For sample D, difference of errors, $d = E_D(h_1) - E_D(h_2)$
For independent samples S_1 and S_2 with size ≥ 30,
Difference of errors, $d' = E_{S1}(h_1) - E_{S2}(h_2)$
Expectation, $E(d') = d'$
Variance, $\sigma_{d'}^{\ 2} = (E_{S1}(h_1)\ (1 - E_{S1}(h_1))\ /\ n_1) + (E_{S2}(h_2)\ (1 - E_{S2}(h_2))\ /\ n_2)$

Bayesian decision theory and learning

Random process: Data generated is not completely known. Observable probability for data x as an outcome of random variable X.

Bayesian Interface: Joint and Conditional probabilities P(A and B) = P(A)P(B|A) = P(B)P(A|B)
Bayes' Theorem: $P(h|D) = P(h)\ P(D|h) / P(D)$, where h is hypothesis and D is data;
P(h|D) -> Posterior probability and P(h) -> Prior probability

Maximum a Posteriori (MAP) Hypothesis: Maximizes posterior probability
$h_{MAP} \equiv argmax_{h \in H}\ P(h|D)$ or $h_{MAP} \equiv argmax_{h \in H}\ P(h)\ P(D|h)$

Maximum Likelihood (ML) Hypothesis: Maximizes likelihood while ignoring prior information
$h_{ML} \equiv argmax_{h \in H}\ P(D|h)$

Features of Bayesian Learning:
- Flexible and dynamic
- Uses prior knowledge of hypothesis
- Uses weight of hypothesis in making decisions
- Provides framework for optimal decision-making

Concept Learning under Bayesian framework:
Likelihood, P(D|h) = 1 if h is element of version space $VS_{H,D}$, otherwise 0
Prior, $P(h) = 1 / |H|$
Marginal probability of data, $P(d) = |VS_{H,D}| / |H|$

Least Mean Square Error estimate as ML hypothesis:
Target function: $y = f(x)$
Hypothesis: h; $y_i = h(x_i) + e_i$ and $e_i \sim N(0, \sigma)$
Mean Square Error, $MSE = \Sigma_{i=1\ to\ n}\ (y_i - h(x_i))^2$

$$P(D|h) = \left(\frac{1}{\sqrt{2\pi\sigma}}\right)^n \prod_{i=1}^{n} e^{\frac{-1}{2}\left(\frac{y_i - h(x_i)}{\sigma}\right)^2}$$

$h_{ML} \equiv argmax_{h \in H}\ \Sigma_{i=1\ to\ n}\ (y_i - h(x_i))^2$

Concepts challenging to comprehend:
Understanding confidence interval and k-fold cross-validation were a bit challenging for me and I had to read the book thoroughly to understand how they are working.

Interesting and exciting concepts:
Bayesian decision theory is an intriguing concept as it would focus on the statistical properties of patterns of dataset in probability densities and thus, makes use of conditional probability to classify target class.

Concepts not understood:
None.

Any novel idea of yours out of the lessons:
The N% confidence interval is an interval bounding both sides, i.e., both from above and below. But there might be a case when we just need that the error should be lesser than a certain quantity, and we don't care whatever small the error is. So, to bound just the maximum error, we have to make the confidence interval one-sided. We know that Normal Distribution is symmetric about its mean, so this modification can be achieved by doubling confidence of the initial N% confidence interval.

Difficulty level of the quiz:
Fair.

Was the time given to you for solving the quiz appropriate?
Yes.

Did the quiz questions enhance your understanding of the topics covered?
Yes, the questions covered the whole of the syllabus for this quiz for which an exhaustive understanding of the topics is a must. Especially, the theoretical questions were tricky and brainstorming.
