**Aluno: Ciro B Rosa**

**Número USP: 2320769**

**E-mail:** [**ciro.rosa@alumni.usp.br**](mailto:ciro.rosa@alumni.usp.br)

**Disciplina: MAC0460/5832 - Introdução ao aprendizado de máquina**

**Lista 2 – Respostas**

**1. Comente sobre o diagrama abaixo. O que o diagrama como um todo ilustra e o que cada**

**componente representa?**



Considere um determinado processo ou sistema, cujas entradas (X) e saídas (y) são descritas exatamente pela função y = f(X). Considere também que a obtenção da função f exata apresenta dificuldades de ordem teórica e/ou prática. O diagrama apresentado resume um método genérico de obtenção de uma função g(X) que se aproxima da função f(X), a partir de:

* Um conjunto de hipóteses H, contendo uma família de funções hm(X) (n=1, ..., M).
* Um algoritmo de aprendizagem A, cujo propósito seria o de escolher uma das funções hm (X) = g(X) (hipótese final) que melhor se aproxima de f(X).

Deste cenário, surge a questão: como avaliar o quanto uma dada função conhecida hm (X) se aproxima de uma função desconhecida f(X)?

Para responder à pergunta, o diagrama propõe o uso de uma base de dados finita (training examples), contendo exemplos de entradas X e saídas y = f(X), como base para escolha da função aproximada g(X) dentre as diversas funções hm (X) contidas no conjunto de hipóteses H.

De modo a garantir comportamento aleatório da base de dados, de forma similar à aleatoriedade observada na população que lhe deu origem, o bloco em vermelho indica a necessidade de aleatoriedade na escolha dos exemplos (X, y). Desta forma, utiliza-se a distribuição estatística da amostra como inferência aproximada da distribuição real de toda a população.

**2. O que é Ein e Eout?**

A letra E denota uma medida de erro entre as respostas reais y observadas e y\_hat previstas em uma população, para um mesmo conjunto X de entradas. Eout refere-se ao erro “out-of-sample”, ou seja, observações reais y e previsões y\_hat para toda a população. Ein refere-se ao erro “in-sample” restrita a uma amostra retirada da população, entre as respostas reais y e as previstas y\_hat.

**3. Quando consideramos a formulação teórica de aprendizado de máquina, uma das possibilidades é investigar o valor |Ein – Eout|. O que esse valor expressa e por que nos interessa investigar ele?**

O valor |Ein – Eout| é uma estimativa do erro que se comete ao estimar a distribuição de probabilidades out-of-sample pela distribuição in-sample. Na formulação teórica, Eout é desconhecido, essencialmente por dois fatores:

* Pelo fato de que a função exata f que relaciona y = f(X) ser desconhecida;
* Pelo fato de que talvez seja inviável na prática conhecer todas as respostas reais y em uma grande população.

Por estes motivos, busca-se estimar Eout através de sua aproximação Ein. Quanto melhor for a similaridade entre essas duas grandezas, mais precisa será a estimativa g(X) da função f(X).

**4. A desigualdade de Hoeffding, no contexto de aprendizado de máquina, com respeito a uma certa hipótese h, é dada por:**

**Explique o significado dessa desigualdade.**

O significado e importância da desigualdade de Hoeffding está na estimativa por um limite superior de uma probabilidade essencialmente desconhecida (o lado esquerdo da desigualdade) por uma quantidade totalmente conhecida e ajustável pelo projeto (lado direito), já que (erro máximo permitido para a aproximação de f(X) por h(X) ) e N (número de observações da amostra) são parâmetros de controle do projeto.

**5. A desigualdade de Hoeffding, no contexto de aprendizado de máquina, quando selecionamos uma hipótese de um espaço com M hipóteses é dada por:**

**Comente sobre a diferença entre essa desigualdade e a do item anterior.**

A questão anterior indica a desigualdade de Hoeffding para o caso de se avaliar o erro de uma única hipótese h, contida no espaço H de hipóteses.

Considerando que o espaço amostral H pode ser escolhido contendo M hipóteses diferentes, a inequação traduz o erro máximo que este conjunto de M hipóteses traz para a solução g(X). Caso as soluções possíveis sejam disjuntas entre si, temos a igualdade da inequação. Caso as soluções possuam algum tipo de sobreposição, temos o “menor que”.

**6. O bound no item anterior foi obtido aplicando-se o union-bound. O que é union-bound?**

Como explicado, a desigualdade de Hoeffding na questão 4 é aplicada para uma única hipótese h, conhecida, escolhida dentro do espaço de hipóteses H. Caso a hipótese g (aproximação de f) fosse conhecida a priori, a desigualdade seria aplicável sem restrições.

O fato é que f(X) é desconhecida e g(X) só será escolhida a posteriori, dentre as opções h(X). O union-bound é a forma de se estimar o erro da função g(X) a priori, a partir da união entre todas as hipóteses h contidas em H.

**7. O que são dicotomias? O que é growth-function? O que é break point? Qual a relação entre eles?**

Seja X um vetor de dados de entrada e seja h uma hipótese que classifica de forma binária o vetor X através de y = h(X). Ou seja, X pode pertencer ao grupamento y = +1 ou y = -1. Nesta descrição, os grupamentos são formalmente conhecidos como dicotomias.

Seja um espaço de hipóteses H, possuindo um número M de hipóteses. De forma geral, M pode ser infinito. Ainda de forma geral, sejam duas hipóteses distintas h1 e h2 pertencentes a H, porém classificando exatamente da mesma forma um mesmo conjunto finito X de entrada. Nesse sentido, h1 e h2 são hipóteses que se equivalem, e podem ser representadas por uma única hipótese “equivalente”. O número total de diferentes hipóteses “equivalentes” de um espaço H é dado pela growth-function mH(N), sendo que mH(N) < M.

Por definição, se nenhum dataset de tamanho k pode ser dividido/classificado pelo espaço de hipóteses H, então k é definido como um break point para H.

**8. O que você entendeu sobre o processo envolvido na troca do M em pelo growth-function mH(N)? Qual o interesse em se fazer essa troca? Qual é o novo bound obtido após a troca?**

Conforme mencionado na questão 7, o número de hipóteses M de um espaço H pode ser infinito, o que leva a um bound infinito para a desigualdade de Hoeffding. A growth function mH(N), por ser um número finito, retoma a desigualdade para um bound também finito.

Considerando-se a growth-function, o bound passa a ser o termo da raiz quadrada abaixo:

**9. Dissemos que a VC dimension relaciona-se com a expressividade do espaço de hipóteses. Comente sobre isso.**

Não me recordo sobre comentário sobre expressividade do espaço de hipóteses, porém tomando como “d” o número de dimensões do vetor de entrada X, a VC-dimension (dVC) de um perceptron seria igual a “d+1”. dVC seria, de certa forma, o tamanho equivalente (menor que d) capaz de dividir o espaço de pontos X.

**10. Como o VC bound é expresso em termos da VC dimension?**

Expressando Eout(g) como sendo:

Temos o bound para Eout definido como em função de dVC:

**11. Baseado no VC bound, explique como podemos calcular o número de amostras necessárias para se garantir uma certa precisão, com probabilidade 1 -δ, supondo que o espaço de hipóteses considerado tem dimensão VC igual a dVC?**

A estimativa do tamanho de amostras N pode ser derivada de forma iterativa, a partir da inequação

onde as quantidades dVC, (erro admitido para o modelo) e (assertividade do modelo) são conhecidos:

* Estima-se um valor inicial para N no lado direito da inequação;
* Calcula-se o lado direito da inequação;
* Se o valor inicial de N é maior que o valor calculado acima, toma-se este como o tamanho da amostra. Caso negativo, repete-se o ciclo com o valor calculado.

**12. Por que apenas garantir não é suficiente?**

Creio que falta a pergunta especificar o “propósito da suficiência” à qual se faz referência. De forma genérica, é possível especular que a expressão acima refere-se à avaliação de uma única hipótese h, a qual não é necessariamente a melhor hipótese do espaço H, denotada por g.

**13. Quais as similaridades e diferenças entre o VC analysis e o Bias-variance analysis?**

**14. Escreva a sua opinião sobre quão úteis são os conteúdos cobertos nas lectures mencionadas para o entendimento sobre Machine Learning.**

Em meu caso específico, tive um treinamento profundo, porém focado apenas na prática, sobre codificação em R para obtenção de algoritmos de ML. As aulas, e em especial a leitura do livro após discussão em classe, trouxe o embasamento teórico que me faltava.