Machine Learning
CPS 863
Terceiro Trimestre de 2024

Professor: Edmundo de Souza e Silva

Lista de Exercícios: 1 (parte b) (exercício feito em classe)

ATENÇÃO! Faça as listas de forma que TODAS AS RESPOSTAS sejam DEVIDAMENTE COMENTADAS (passos para se chegar a resposta).

Questão 1

Este exercício foi feito em classe no dia 10/Out/2024. A lista contém as questões resolvidas e ainda alguns itens a mais. Faça a lista complete as questões que faltaram.

Este exercício é motivado pelo trabalho em `https://ieeexplore.ieee.org/document/9006548`, Seção H (*Leveraging spatio-temporal correlation across homes*). O problema foi simplificado neste exercício.

Imagine que dispomos de um classificador implementado em roteadores residenciais de um provedor de Internet (ISP). A cada janela de tempo (por exemplo a cada 5 minutos) o classificador do roteador i fornece como saída uma dentre 2 possibilidades: (a) existe um ataque DDoS acontecendo a partir da residência do roteador i, nesta janela de tempo; (b) não há ataque acontecendo a partir da residência do roteador i nesta janela.

A cada 5 minutos o ISP amostra o resultado de M roteadores escolhidos de forma aleatória dentre todos os roteadores da sua base que, para todos os efeitos deste problema, pode ser considerada como muito grande (infinita). O objetivo do ISP é determinar, a partir das M amostras coletadas, se um ataque aconteceu ou não durante a janela de tempo amostrada. Em outras palavras, o ISP quer determinar a possibilidade de uma das seguintes hipóteses serem verdadeiras: h_a (há um ataque DDoS acontecendo na rede do ISP na janela amostrada) ou h_b que é a hipótese complementar.

O ISP conhece o classificador usado em cada roteador residencial, e sabe que o resultados não é 100% confiável. Portanto, ele usará correlação espacial conforme sugerido no artigo acima e explicado em classe.

No que se segue usaremos algumas definições comuns que podem ser encontradas em `https://en.wikipedia.org/wiki/Confusion_matrix` (ver também a figura em `https://en.wikipedia.org/wiki/Confusion_matrix`).

Notação:

- M: número de roteadores amostrados (em uma janela de tempo);
- Inf: variável aleatória (va) indicando se a residência é um "bot", isto é, está ou não infectada. $P[\text{Inf}]$ ($P[\overline{\text{Inf}}]$) é a probabilidade de uma residência estar infectada (não estar infectada);
- TPR: (true positive rate ou hit rate) taxa de acerto do classificador, ou probabilidade do classificador corretamente sinalizar um ataque, dado que um ataque está acontecendo no ISP (Nota: obviamente somente residências infectadas podem gerar um ataque quando ele ocorre);
- FPR: (false positive rate) ou probabilidade do classificador do roteador residencial erradamente sinalizar um ataque a partir da residência;
- L: variável aleatória indicadora $L = 1$ se o roteador alarma, $L = 0$, caso contrário.
- $P[h_a]$ probabilidade de ocorrer um ataque DDoS no ISP em uma janela de tempo. (Se você tem algum conhecimento prévio sobre ataques, talvez possa estimar o $P[h_a]$).

Suponha que, em uma determinada janela de tempo, das M amostras coletadas, V roteadores sinalizaram que um ataque estava ocorrendo na janela (e então $M - V$ roteadores sinalizaram que tudo estava normal nas suas respectivas residências). Suponha ainda que um ataque ocorre em um intervalo independentemente das infecções nas residências.

Para os seus cálculos, suponha que: TPR $= 0.8$, FPR $= 0.1$, $P[\text{Inf}] = 0.2$. Como você não tem conhecimento prévio sobre $P[h_a]$, suponha inicialmente que $P[h_a] = P[h_b] = 0.5$, $V = 20$, $M = 200$.

Responda as seguintes perguntas, mas só substitua os valores no final:

1. Suponha que um ataque esteja ocorrendo.
   Calcule $P[L = 1|\text{Inf}, h_a]$ e $P[L = 1|\overline{\text{Inf}}, h_a]$ e então $P[L = 1|h_a]$ e $P[L = 0|h_a]$.

2. Suponha que um ataque **não** esteja ocorrendo.
   Calcule $P[L = 1|\text{Inf}, h_b]$ e $P[L = 1|\overline{\text{Inf}}, h_b]$, e então $P[L = 1|h_b]$ e $P[L = 0|h_b]$.

3. Calcule $P[\mathcal{D}|h_a]$ em função de V e M. (Likelihood)

4. Calcule $P[\mathcal{D}|h_b]$ em função de V e M (Likelihood)

5. Calcule $P[h_a|\mathcal{D}]$ e $P[h_b|\mathcal{D}]$ (Posterior)

6. Qual o mínimo de roteadores que deveriam alarmar (V) para que você tenha confiança que um ataque ocorreu.
7. Caso $P[h_a] = 0.1$ os resultados variam?
   Trace as curvas $P[h_a|\mathcal{D}]$ e $P[h_b|\mathcal{D}]$ em função de V e explique as curvas.
8. Caso $P[h_a] = 0.1$
   Trace a curva $\log(P[h_a|\mathcal{D}]/P[h_b|\mathcal{D}])$ em função de V.
9. Para TPR $= 0.9$ e FPR $= 0.1$, plote, em um mesmo gráfico, a função de probabilidade de massa: (a) do número de roteadores que alarmam quando há um ataque; (b) do número de roteadores que alarmam quando **não** há um ataque. Na implementação do classificador central (aquele que recebe os sinais dos roteadores domésticos, e que são os "sensores" em cada residência), você deve decidir a partir de quantos roteadores residenciais alarmando o classificador central deverá detectar que um ataque estã ocorrendo.
   (a) Explique como avaliar o erro da sua decisão.
   (b) Estime esse erro para o valor escolhido.
10. O classificador central comete erros, evidentemente.
    (a) Calcula o TPR_c e o FPR_c do classificador central.
    (b) Plote a *ROC curve* do classificador central.
    (c) Compare o classificador central com o classificador residencial.