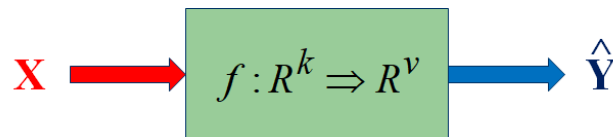


Modelagem e Interpretação da Resposta do Modelo

1. Supondo um problema de Classificação (Classe A e Classe B) e você tenha modelado a solução com $k = 10$ e $v = 1$. Como você interpretaria a resposta do modelo em relação às duas possíveis classes?
2. O que você faria caso desejasse minimizar concomitantemente as ocorrências de FP e FN?
3. É viável definir uma interpretação que minimize a ocorrência apenas de FP ou FN? Como?
4. Caso na modelagem o projetista tenha optado por representar a solução do problema com $v = 2$, Como você interpretaria a resposta como sendo Classe A ou B?
5. Uma boa prática, em problemas de classificação, é modelar o vetor de saída com uma dimensão igual ao número de classes. Nesses casos, como são definidas as respostas para cada classe? Que característica importante possui essa forma de representação?
6. Suponha que você após o treino você aplique o modelo sobre o próprio conjunto de treino e obtenha como resposta a seguinte matriz de confusão:

Estimado \ Desejado	A	B	C	D	E
A	70	5	0	3	1
B	6	65	8	0	2
C	3	10	75	1	0
D	1	3	2	95	2
E	5	0	1	1	100

- a. Qual o tamanho do conjunto de treino e o número de exemplos de cada classe?
- b. O treino usa as classes de forma balanceada?
- c. Qual a acurácia alcançada no treino?
- d. Qual a classe melhor aprendida pelo modelo?
- e. Considerando a classe C, qual a taxa de FP e de FN?
- f. Qual a precisão no reconhecimento da classe A?

Exercício Prático

A partir do arquivo de dados em anexo – DL_Testes02, cujo problema é o diagnóstico do câncer de mama, realize as operações a seguir e forneça os resultados obtidos:

1. Verifique os dados do problema e defina quem você quer considerar diagnóstico positivo e quem considerar negativo.
2. Considerando o número de exemplos disponíveis, que estratégia você usaria para definir os conjuntos de treino, validação e teste?
3. Fazendo de contas que a planilha contém apenas dados de teste e que a resposta do modelo foi a constante na coluna mais à direita, calcule a Matriz de Confusão, a acurácia e as taxas de falso positivo e falso negativo alcançadas.
4. Monte o gráfico e identifique, se possível, o valor do limiar de forma a reduzir, concomitantemente, os FPs e FNs.
5. Caso você precise interpretar a saída de forma a reduzir a ocorrência de FNs (diagnostica como boa uma pessoa com câncer), o que você faria? Mostre como ficaria a nova Matriz de Confusão, analise e explique as mudanças ocorridas.