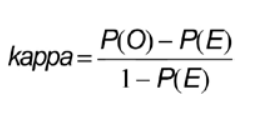
**Relatório Individual - Aula Classificação**

**Aluno: Clésio de Araújo Gonçalves**

Para todos os exercícios propostos, considere as seguintes definições:

* **Acurácia**: A acurácia representa a porcentagem de casos classificados corretamente no conjunto de testes. A acurácia apresenta melhores resultados para bases balanceadas.
* **Índice kappa:** mostra a concordância dos resultados obtidos e fornece uma ideia de quanto eles se afastam ou se aproximam do esperado. Os valores do índice kappa podem ser interpretados assim: < 0,4 (pobre) e > 0,75 (excelente). Quanto mais próximo de zero for o valor de kappa, maior a discordância dos dados.



Onde:

P(O): proporção observada de concordâncias (soma das respostas concordantes dividida pelo total);

P(E): proporção esperada de concordâncias (soma dos valores esperados das respostas concordantes dividida pelo total).

**Exercício 1: Treine uma random forest e calcule os resultados de Acurácia e Kappa para o conjunto de treino/teste**

Realizei o treinamento do dataset utilizando o classificador RandomForestClassifier e obtive os seguintes resultados:

1. **Conjunto de treino**
   * Acurácia treino: 1.0
   * Kappa treino: 1.0
   * Matriz de confusão treino:

**[[35 0 0]**  → classe 0

**[ 0 35 0]**  → classe 1

**[ 0 0 35]]**  → classe 2

O conjunto de dados de treino apresenta um total de 35 amostras para cada classe (classe 0, classe 1 e classe 2). O classificador realizou a predição com 35 amostras corretas para a classe 0; 35 amostras corretas para a classe 1; e 35 amostras corretas para a classe 2. Como a classificação obteve 100% de acerto, a acurácia e o índice kappa obtiveram valor de 1.0

1. **Conjunto de teste**
   * Acurácia teste: 0.9555555555555556
   * Kappa teste: 0.9333333333333333
   * Matriz de confusão teste:

**[[15 0 0]** → classe 0

**[ 0 14 1]** → classe 1

**[ 0 1 14]]** → classe 2

O conjunto de dados de teste apresenta um total de 15 amostras para cada classe (classe 0, classe 1 e classe 2). O classificador realizou a predição com 15 amostras corretas para a classe 0; na classe 1, 14 amostras foram identificadas corretamente e 1 amostra incorretamente classificada como classe 2; e na classe 2, 14 amostras foram corretamente classificada como classe 2 e 1 amostra incorretamente classificada como da classe 1.

A acurácia e o índice kappa são excelentes. A acurácia obteve 43 acertos de um total de 45 amostras, totalizando 95% de acerto. O índice kappa obteve um excelente resultado (93%), indicando baixa discordância dos dados.

**Os resultados são excelentes para ambos datasets de treino e teste, devido ao fato desses conjuntos de dados apresentarem classes balanceadas.**

**Exercício 2: Compare os resultados obtidos pela MLP treinada com os dados normalizados e não normalizados**

Cabe destacar que o conjunto de dados de teste possui um total de 150 amostras para a classe 0 e 81 amostras para a classe 1. Ou seja, apresenta dados desbalanceados.

1. **O conjunto de dados de teste NÃO NORMALIZADO apresenta como resultado:**

Acurácia = 0.6883116883116883

Kappa = 0.2786885245901639

**[[123 27]** → classe 0

**[ 45 36]]** → classe 1

Para a classe 0, foram classificadas corretamente 123 amostras e 27 amostras foram classificadas incorretamente - como sendo da classe 1. Para a classe 1, foram classificadas corretamente 36 amostras e 45 amostras foram classificadas incorretamente - como sendo da classe 0.

A acurácia apresentou um resultado relativamente alto (68%), pois acertou 159 amostras de um total de 231. Como a base está desbalanceada, esse resultado não condiz com a realidade, pois para a classe 1, há mais amostras classificadas incorretamente do que amostras classificadas corretamente.

O índice kappa apresenta péssimo resultado (27%), com valor menor que 40%. Indicando alta discordância dos dados.

1. **O conjunto de dados de teste NORMALIZADO apresenta como resultado:**

Acurácia Dados Normalizados: 0.6796536796536796

Kappa Dados Normalizados: 0.28846153846153844

**[[115 35]** → classe 0

**[ 39 42]]** → classe 1

Para a classe 0, foram classificadas corretamente 115 amostras e 35 amostras foram classificadas incorretamente - como sendo da classe 1. Para a classe 1, foram classificadas corretamente 42 amostras e 39 amostras foram classificadas incorretamente - como sendo da classe 0.

A acurácia apresentou um resultado um pouco mais baixo (mas similar) do que o resultado com o conjunto de dados não normalizados, pois acertou 157 amostras de um total de 231. Como a base está desbalanceada, esse resultado não condiz com a realidade, pois para a classe 1, ainda há várias amostras classificadas incorretamente.

O índice kappa também apresenta um resultado similar (28%), ligeiramente mais alto. Resultado este ainda considerado péssimo, com valor menor que 40%. Indicando alta discordância dos dados.

**Exercício 3: Faça o balanceamento dos dados de treino replicando os exemplos da classe minoritária. Em seguida, treine novamente a MLP com os dados normalizados e observe o que acontece com os resultados.**

O conjunto de dados de teste possui um total 231 amostras, sendo dividido em 150 amostras para a classe 0 e 81 amostras para a classe 1. Ou seja, apresenta dados desbalanceados.

O conjunto de dados de treino possui um total de 537 amostras, sendo dividido em 350 amostras para a classe 0 e 187 amostras para a classe 1. Ou seja, apresenta dados desbalanceados.

Com isso, repliquei os dados da classe 1 somente no dataset de treino, obtendo 374 amostras (187 \* 2) para a classe 1. A classe 0 continua com 350 amostras. Com isso, o conjunto de dados de treino está balanceado, agora totalizando 724 amostras.

Cabe destacar que nenhuma alteração foi feita no dataset de teste, continuando com 231 amostras divididas entre a classe 0 e a classe 1.

Após o balanceamento dos dados, obtive os seguintes resultados (com os dados normalizados):

1. **Conjunto de dados de treino:**

**Acurácia treino: 0.9668508287292817**

**Kappa treino: 0.933506811572019**

**[[328 22]** → classe 0

**[ 2 372]]** → classe 1

Para a classe 0, foram classificadas corretamente 328 amostras e 22 amostras foram classificadas incorretamente - como sendo da classe 1. Para a classe 1, foram classificadas corretamente 372 amostras e 2 amostras foram classificadas incorretamente - como sendo da classe 0.

A acurácia apresentou um excelente resultado (96%) com os dados balanceados, classificando corretamente 700 amostras de um total de 724 amostras. Esse resultado é muito mais alto comparado aos 74% obtidos com os dados desbalanceados.

O índice kappa também apresenta um excelente resultado (93%) com os dados balanceados, indicando baixa discordância dos dados. Esse resultado é muito mais alto comparado aos 39% obtidos com os dados desbalanceados.

**Assim, pode-se concluir que a acurácia e o índice kappa apresentam melhores resultados quando a base possui as classes balanceadas.**

1. **Conjunto de dados de teste:**

Acurácia teste: 0.7012987012987013

Kappa teste: 0.3496143958868895

**[[114 36]** → classe 0

**[ 33 48]]** → classe 1

Para a classe 0, foram classificadas corretamente 114 amostras e 36 amostras foram classificadas incorretamente - como sendo da classe 1. Para a classe 1, foram classificadas corretamente 48 amostras e 33 amostras foram classificadas incorretamente - como sendo da classe 0.

A acurácia apresentou um resultado de 70% com os dados desbalanceados, classificando corretamente 162 amostras de um total de 231 amostras. Esse resultado é similar aos 67% obtidos anteriormente no exercício 2.

O índice kappa apresenta um resultado de 34% com os dados desbalanceados, indicando alta discordância dos dados. Esse resultado também é similar aos 28% obtidos anteriormente no exercício 2.

**Os resultados da acurácia e índice kappa estão muito abaixo dos resultados obtidos no conjunto de dados de treino com as classes balanceadas.**