Primeiro transformei os dados para que correspondesse ao (X, y) do conjunto de treino para serem aplicados mais facilmente os algoritmos. Depois apliquei os dados para separação das bases de treino e validação com 30% dos dados para teste, em um caso com estado aleatório definido e outro automático, em uma mlp com ativação relu, o que causou uma redução da assertividade de 60% para 40% respectivamente. Obteve resultados melhores com a ativação definida como indentity, com resultados de até a 80%.

Na aplicação do K-fold foram aplicados duas quantidades para medias dos resultados, uma definida com 10 e outra que corresponde ao tamanho da base de testes. Com a média definida pelo tamanho da base foi possível obter alguns resultados melhores em algumas execuções, por exemplo, definido como 10 dificilmente passou de 50%, pelo tamanho da base alcançou 60%. Essa mudança não é muito significativa, por conta de a base de dados não apresentar também um tamanho mais elevado, já que o cálculo é limitado pelo tamanho da base.

Nas métricas com uma aplicação de uma mlp, foi possível observar que o algoritmo foi bom tanto para acertar positivos quanto acerto de negativos, se mantendo acima de 58% na maioria das execuções, em algumas foi possível observar uma prevalência no acerto de positivos sob o acerto de negativos.

Em relação a redes neurais os valores resultantes foram os mesmos na aplicação das quatro variações ("identity", "logistic", "tanh" e "relu"), se mantendo linear na maioria das execuções, o que se justifica pelo tamanho do número de elementos de treino, que é 9. Ao diminuir esse valor, foi possível obter resultados diferentes, abaixo segue os melhores resultados:

* Número de splits = 8 -----> resultados variáveis abaixo de 31,25%. Rodando o código novamente para uma reorganização da base apresentou resultados de 68,75%.
* Número de splits = 7 -----> apenas logistic com 68,28% e as outras com 71,28%.
* Número de splits = 6 -----> resultados variáveis em todas as classes, hora 58,33% hora 66,66%.
* Número de splits = 5 -----> todos os resultados variando hora 60% hora 70%.
* Número de splits = 4 -----> resultados variáveis em todas as classes, apresentando resultados como 37.5%, 66,66% e 70,83%.

Com numero de splits menores os resultados diminuirão ainda mais. Por fim, fiz um balanceamento na base de estudo, no entanto os resultados apresentados para todos os algoritmos foram menores que 50% em sua maioria.

Aluno: Ivo Ireneu de Souza Juniors

Professor: Victor Lundgren

Disciplina: TAIA-SI-UAST