



Universidade de Brasília

DEPARTAMENTO DE ESTATÍSTICA

18 de junho de 2024

Lista 4: Classificação de imagens.

Prof. Guilherme Rodrigues

Introdução às Redes Neurais Profundas

Tópicos especiais em Estatística 1

- (A) O aluno poderá consultar materiais relevantes disponíveis na internet, tais como livros, *blogs* e artigos.
- (B) O trabalho é individual ou em duplas.
- (C) O aluno deverá enviar o trabalho até a data especificada na plataforma Microsoft Teams.
- (D) O trabalho será avaliado considerando o nível de qualidade do relatório, o que inclui a pertinência das soluções encontradas, e o desempenho do modelo, medido pela estatística *F1-score*.
- (E) Escreva seu código com esmero, evitando operações redundantes, comentando os resultados e usando as melhores práticas em programação.

“Datiloscopia ou papiloscopia é o processo de identificação humana por meio das impressões digitais, normalmente utilizado para fins judiciais. Esta área do conhecimento estuda as papilas dérmicas (saliências da pele) existentes na palma das mãos e na planta dos pés, também conhecida como o estudo das impressões digitais.” (*Wikipedia*)

Neste trabalho, ajustaremos redes neurais convolucionais (*CNN*) para classificar impressões digitais de acordo com o sexo biológico do indivíduo. Nosso objetivo, portanto, é maximizar a probabilidade de acerto do modelo em imagens não utilizadas no processo de treinamento.

Entre outras aplicações, tal ferramenta poderia auxiliar o trabalho investigativo da Polícia Federal em situações criminais, afunilando o universo de suspeitos a partir da identificação do sexo vinculado a um fragmento papilar coletado na cena do crime. De maneira similar, em desastres de massa, a classificação do sexo dos restos mortais contribuiria para o processo de identificação dos corpos das vítimas.

O banco de imagens que utilizaremos (disponível em https://1drv.ms/f/s!Aiqt4i7wEOAe24dzk9tyQQx9R_8z3w?e=iclyFp) contém impressões digitais dos dedos das mãos de 400 pessoas. Os nomes dos arquivos do conjunto de teste estão no formato “Sexo_Indice” (por exemplo, “F_0001” corresponde a primeira foto de uma pessoa do sexo feminino). Para viabilizar o exercício, o sexo das imagens do conjunto de teste não foi disponibilizado.

Use as ferramentas disponíveis para melhorar o desempenho do modelo, incluindo, entre outras, a **criação de dados sintéticos**, computação em nuvem, *dropout*, *batch normalization*, *boosting*, *bagging* e *autoML*.

Para completar a atividade, solicito o envio dos arquivos a seguir.

- 1) Relatório em *pdf* ou *html* descrevendo a metodologia adotada (com o respectivo código fonte); Sugiro incluir um pequeno quadro indicando as principais soluções adotadas.
- 2) Arquivo em formato *.csv* contendo duas colunas: “Nome”: Nome dos arquivos do banco de teste; “Classificação”: Classificação atribuída pelo modelo construído (“M” ou “F”).