

Projeto Prático 02

Web services são frequentemente protegidos por desafios que são fáceis para humanos mas de difícil resolução automática por computadores. Isso ocorre por diferentes propósitos, incluindo redução de spams e prevenção de ataques a senhas usando força-bruta. Tais desafios são denominados CAPTCHA (*Completely Automated Public Turing test to tell Computers and Humans Apart*) ou HIP (*Human Interactive Proof*).

Asirra (*Animal Species Image Recognition for Restricting Access*) é um HIP que solicita que os usuários diferenciem imagens de gatos e cachorros. Embora a base de dados contemple imagens com diferentes ângulos, iluminação e fundos, a literatura recente já relata classificadores com taxas de acerto acima de 80%. Com isso, o método não é mais considerado seguro (para um classificador com 60% de acurácia, a probabilidade de acertar um HIP com 12 imagens é de 1/459, contra 1/4096 da abordagem aleatória).

Uma maneira de dificultar o reconhecimento automático seria utilizar mais categorias de objetos e requisitar ao usuário para escolher dentre os possíveis candidatos quais deles pertencem a determinada categoria.

Tarefa

Neste projeto, o objetivo é utilizar os algoritmos vistos (e que serão vistos) na disciplina para classificar imagens dos seguintes problemas:

- Dogs vs. Cats (classificação binária): O problema e a base de dados estão disponíveis na plataforma *Kaggle*¹.
- Cifar 10 (classificação multiclasse): o desafio consiste em diferenciar 10 classes de objetos: avião, automóvel, passarinho, gato, cachorro, veado, sapo, cavalo, barco e caminhão. O problema e a base de dados estão disponíveis na plataforma *Kaggle*².

Objetivos

- Inicialmente, teste o arquivo `projeto02.py`. Nele, já estão implementados, especificamente para o problema **Dogs vs. Cats**: (a) uma opção para acesso às imagens de treinamento, (b) dois descritores simples e (c) a chamada a diferentes algoritmos de classificação. Adapte ou crie um código próprio para o problema **Cifar10**.
- Sua tarefa consiste em identificar: (a) as características mais relevantes (ou seja, que conduzem a melhores taxas de classificação) e (b) os melhores parâmetros para os algoritmos de classificação.
- Discuta a dificuldade encontrada ao resolver um problema de classificação binária em comparação àquela encontrada em um problema multiclasse.
- Não se esqueça que não é apenas o resultado final que importa. Salve os resultados intermediários de cada etapa, visando justificar as suas escolhas.

¹Links:

– <https://www.kaggle.com/c/dogs-vs-cats>
– <https://www.kaggle.com/c/dogs-vs-cats-redux-kernels-edition>

²<https://www.kaggle.com/c/cifar-10>

Entrega

O trabalho deve ser desenvolvido em equipes de **até três pessoas**. Cada equipe deve submeter (via moodle) um arquivo .zip com:

- Os arquivos Python com as implementações realizadas. Não se esqueça de comentar adequadamente o código.
- Um arquivo pdf com um breve relatório.