

Projetos 2024/25 — OML (DMAT, UM)

1 Sobre os projetos

1.1 Objetivo

O objetivo dos projetos é realizar um trabalho sobre a análise e implementação de um tema abordado nas aulas com uma parte de programação e de experiência numérica.

- Serão criados 4 grupos de 4 alunos;
- No final deve-se produzir um relatório de aproximadamente 15 páginas (eventuais anexos não contam para este limite) e um conjunto de slides para apresentar publicamente o trabalho.
- A classificação deste trabalho será: 50% do relatório, 50% da apresentação e discussão oral.

1.2 Procedimentos

- **Quinta-feira, 10/04, a partir de 17h00**, cada coordenador de grupo envia um email para `gjm@math.uminho.pt` e `sofialopes@math.uminho.pt` com a constituição do grupo e com a ordem de preferência dos quatro projetos. Vão-se atribuir os projetos consoante a hora de chegada dos emails (excluindo os emails que chegarem mais cedo).
- Após terem conhecimento do projeto que vão efetivamente realizar, o representante do grupo deve contactar o orientador para a organização de sessões de trabalho.
- O representante deve enviar por email, os pdf do relatório e da apresentação ao seu orientador às 18h00 do dia 30 maio.
- As apresentações serão no dia 3 de junho, das 15:00 às 19:00 (20 minutos para a apresentação e 30 minutos para discussão por grupo).

2 Temas

T1: Classificador logístico multiclasse: OvA vs ECOC [Orientador GJM]

Descrição. Propõe-se neste trabalho comparar (tempo computacional e métricas de qualidade) as técnicas *one-vs-all* (OvA) e *Error-Correcting Output Codes* (ECOC) do classificador logístico multi-classe nas versões primal e dual com *kernel* polinomial. Deve-se testar diferentes regras de aprendizagem. Deve-se também considerar a versão *mini-batch* com diferentes tamanhos (desde um até ao tamanho da base de dados) e uma versão “*mini-batch*” com tamanho um mas em que a base de dados é percorrida sequencialmente. Inicialmente deve-se considerar bases de dados sintéticas com três classes com características diferentes mas apenas com dois atributos ($I = 2$) por forma a se poder fazer a visualização dos resultados. Deve-se depois considerar uma base de dados real com pelo menos quatro classes. Valorizam-se estratégias de otimização do código. O grupo pode escolher a linguagem de programação que vai usar.

Trabalhos a realizar.

1. Descrição e implementação do método primal com as versões OvA e ECOC.
2. Validação e avaliação com as bases de dados.
3. Descrição e implementação do método dual com *kernel* polinomial com as versões OvA e ECOC.
4. Validação e avaliação com as base de dados.

Os documentos de referências são http://ciml.info/dl/v0_99/ciml-v0_99-all.pdf, nomeadamente a secção 6.2, e https://www.ccs.neu.edu/home/vip/teach/MLcourse/4_boosting/lecture_notes/ecoc/ecoc.pdf.

T2: Classificador logístico multiclasse: OvO vs ECOC [Orientador GJM]

Descrição. Propõe-se neste trabalho comparar (tempo computacional e métricas de qualidade) as técnicas *one-vs-one* (OvO) e *Error-Correcting Output Codes* (ECOC) do classificador logístico multi-classe nas versões primal e dual com *kernel* polinomial. Deve-se testar diferentes regras de aprendizagem. Inicialmente deve-se considerar bases de dados sintéticas com três classes com características diferentes mas apenas com dois atributos ($I = 2$) por forma a se poder fazer a visualização dos resultados. Deve-se depois considerar uma base de dados real com pelo menos quatro classes. Valorizam-se estratégias de otimização do código. O grupo pode escolher a linguagem de programação que vai usar.

Trabalhos a realizar.

1. Descrição e implementação do método primal com as versões OvO e ECOC.
2. Validação e avaliação com as bases de dados.
3. Descrição e implementação do método dual com *kernel* com as versões OvO e ECOC.
4. Validação e avaliação com as bases de dados.

Os documentos de referências são http://ciml.info/dl/v0_99/ciml-v0_99-all.pdf, nomeadamente a secção 6.2, e https://www.ccs.neu.edu/home/vip/teach/MLcourse/4_boosting/lecture_notes/ecoc/ecoc.pdf.

T3: Estimação de parâmetros de um modelo utilizando algoritmos de otimização sem restrições [Orientador SOL]

Descrição. A equação de Horton descreve o comportamento da infiltração da água no solo. Contudo, existe(m) parâmetro(s) que são difíceis de estimar para cada tipo de solo. Recorrendo a dados experimentais e à otimização, pretende-se estimar este(s) valor(es).

Trabalhos a realizar.

1. Descrição da equação de Horton.
2. Descrição do Método do Gradiente, Método de Newton, Método BFGS e Método do Gradiente Conjugado de Fletcher-Reeves.
3. Formulação do problema de minimização em função dos parâmetros desconhecidos.
4. Implementação e comparação dos diferentes métodos com diferentes pontos iniciais e diferentes passos.
5. Interpretação dos resultados obtidos.

Os documentos de referência são <https://link.springer.com/book/10.1007/978-0-387-40065-5>, nomeadamente as secções 2.2, 5.2 e 6.1. Os dados encontram-se no ficheiro OML_projetos_2425-tm3.xlsx.

T4: Estimação de parâmetros de um modelo utilizando algoritmos de otimização sem restrições [Orientador SOL]

Igual ao tema T3 mas considerando os dados do ficheiro OML_projetos_2425-tm4.xlsx.