



Plano de Desenvolvimento da Disciplina

MC846 - Aprendizado de Máquina Supervisionado

Docente: Prof. Marcos M. Raimundo

Descrição

Este documento descreve o plano de desenvolvimento da disciplina (PDD) de MC846 - Aprendizado de Máquina Supervisionado. Em conformidade com o Regimento Geral de Graduação e Pós-Graduação da Unicamp, este plano detalha o programa do curso, a fundamentação teórica dos blocos de conteúdo, o cronograma das avaliações, os critérios de atribuição de nota, a política contra fraudes e plágio e a bibliografia recomendada.

1 Programa da disciplina

O programa desta disciplina aborda os algoritmos de Aprendizado de Máquina Supervisionado por meio de uma perspectiva probabilística, estatística e de otimização contínua. O objetivo terminal é capacitar o aluno a: a) Formular problemas preditivos de classificação e regressão sob uma ótica estatística rigorosa; b) Compreender as garantias de generalização, decomposição de erro e tradeoffs de capacidade dos modelos; c) Derivar analiticamente e implementar algoritmos supervisionados clássicos e modernos, desde estimadores lineares paramétricos até métodos de ensemble e redes neurais profundas.

O curso está estruturado em três blocos fundamentais:

- **Bloco 1: Fundamentos de Probabilidade, Particionamento e Reamostragem**

Objetivo: Estabelecer as bases da inferência Bayesiana, estimação não-paramétrica por espaço de busca e protocolos rigorosos de avaliação de generalização.

Conteúdo: Ajuste de densidades contínuas (Distribuição Beta) e limiares de decisão baseados em taxas de erro (Erros Tipo I e II); distribuições condicionais, Teorema de Bayes e a premissa de independência no Naive Bayes; particionamento guloso não-paramétrico com algoritmos de árvores de decisão (CART), entropia/Gini como incerteza categórica e verossimilhança de perfil; validação de modelos, simulação da distribuição conjunta $P(X, Y)$ via validação cruzada e métodos de Bootstrap.

- **Bloco 2: A História Linear Paramétrica e Inferência Bayesiana**

Objetivo: Desenvolver a teoria dos Modelos Lineares Generalizados (GLMs), métodos de estimação frequentista versus bayesiana e a análise teórica de erros.

Conteúdo: O modelo de ruído Gaussiano homocedástico e a prova de equivalência entre Mínimos Quadrados (OLS) e Máxima Verossimilhança (MLE); Regressão Logística sob a família exponencial (Bernoulli) e derivação da perda de entropia cruzada via função de ligação (*logit*); regularização frequentista (L_1, L_2) e estimação Máxima A Posteriori (MAP) com prioris Gaussianas (Ridge) e de Laplace (Lasso); decomposição formal matemática do erro (Tradeoff Viés-Variância e Erro Irreduzível).

- **Bloco 3: Modelos Avançados, Ensembles e Redes Neurais**

Objetivo: Explorar abordagens de alta capacidade para redução de viés/variância, otimização dual restrita e aprendizado de representação em camadas.

Conteúdo: Teoria de ensembles para redução de variância, amostragem de atributos e aleatorização estrutural (Bagging e Random Forests); otimização numérica no espaço de funções, expansão de Taylor de perdas e pseudo-resíduos (Gradient Boosting e AdaBoost); classificadores de margem máxima (SVMs), formulação dual de Lagrange, *Hinge Loss* e calibração probabilística (Platt scaling); Perceptrons Multicamadas (MLPs) vistos como cadeias de GLMs, funções de ativação finais (Softmax) para distribuições multinomiais e otimização por retropropagação.

2 Atividades e Avaliação

A avaliação da disciplina será composta por três testes teóricos/práticos individuais (T_1, T_2, T_3), aplicados ao término de cada bloco de conteúdo, e um Trabalho Final prático (TF) focado na implementação, validação e análise comparativa de modelos supervisionados.

A nota final (NF) será calculada pela seguinte média ponderada:

$$NF = 0.20T_1 + 0.20T_2 + 0.20T_3 + 0.40TF$$

onde:

- T_1 é a nota do Teste 1 (com foco no Bloco 1), aplicado no dia **14 de setembro**.
- T_2 é a nota do Teste 2 (com foco no Bloco 2), aplicado no dia **26 de outubro**.
- T_3 é a nota do Teste 3 (com foco no Bloco 3), aplicado no dia **30 de novembro**.
- TF é a nota do Trabalho Final (40% do total), a ser entregue no encerramento do semestre.

3 Punição para fraudes e plágios

Detecção de fraude ou plágio em qualquer teste individual ou no trabalho final resultará em nota zero na atividade específica para todos os envolvidos. Casos de reincidência ou plágio grave acarretarão em nota final zero ($NF = 0$) e atribuição de conceito D para todos os envolvidos, com a devida notificação à coordenação do IC/Unicamp.

4 Bibliografia Recomendada

- Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2009.
- James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning: with applications in R*. Springer, 2021.
- Murphy, Kevin P. *Probabilistic machine learning: An introduction*. MIT Press, 2022.
- Bishop, Christopher M. *Pattern recognition and machine learning*. Springer, 2006.