UNIVERSIDADE FEDERAL DE OURO PRETO
PRÓ-REITORIA DE GRADUCAÇÃO
PLANO DE ENSINO

Nome do Componente Curricular em português: **Métodos Estatísticos para Machine Learning** Nome do Componente Curricular em inglês: Statistical Methods for Machine Learning	Código: EST335
Nome e sigla do departamento: Departamento de Estatística DEEST	Unidade acadêmica: ICEB
Nome do docente: Tiago Martins Pereira	

Carga horária semestral	Carga horária semanal teórica	Carga horária semanal prática
60 horas	04 horas/aula	00 horas/aula

Data de aprovação na assembleia departamental: / / 2026

Ementa: Regressão: Introdução. Métodos paramétricos. Métodos paramétricos em altas dimensões. Métodos não paramétricos. Métodos não paramétricos em altas dimensões. Outros aspectos de regressão. Classificação: Introdução. Métodos de classificação. Outros aspectos de classificação. Aprendizado não supervisionado: Redução da dimensionalidade e Mudança de representação. Análise de Agrupamento. Regras de Associação. Sistemas de Recomendação.

Conteúdo programático

Parte I – Regressão:

- Introdução: Notação e suposições. Predição versus Inferência. As duas culturas. A função risco. Seleção de modelos: super e sub-ajuste. Tuning parameters.
- Métodos paramétricos: O método dos mínimos quadrados. Mínimos quadrados quando a suposição de linearidade falha.
- Métodos paramétricos em altas dimensões: Busca pelo melhor subconjunto de covariáveis. Regressão Stepwise. Lasso. Regressão Ridge.
- Métodos não paramétricos: Séries ortogonais. Splines. Knn. Nadaraya-Watson. Regressão polinomial local. Métodos baseados em RKHSs. Arvores de regressão. Bagging e Florestas aleatórias. Boosting. Redes neurais artificiais.
- Métodos não paramétricos em altas dimensões: Taxas de convergência e a maldição da dimensionalidade. Knn e Regressão linear local. Support Vectorc Regression. Florestas aleatórias.
- Outros aspectos de regressão: Interpretabilidade. LIME. PDP e ICE.

Parte II – Classificação:

- Introdução: Função de risco. Estimação do risco e seleção de modelos. Balanço entre Viés e Variância. Outras medidas de desempenho.
- Métodos de classificação: Classificadores Plug-in. Support Vector Machines (SVM). Arvores de classificação. Bagging e Florestas aleatórias. Boosting. Knn. Redes neurais artificiais.

- Outros aspectos de classificação: Assimetria na função de perda, conjuntos desbalanceados e outros cortes. Combinando classificadores.

Parte III – Aprendizado não supervisionado:

- Redução de dimensionalidade e mudança de representação: PCA. Kernel PCA. Projeções aleatórias. Autoencoders.
- Análise de agrupamento: k-médias. Agrupamento espectral. Métodos hierárquicos.
- Regras de associação.
- Sistemas de recomendação: Filtro colaborativo baseado no usuário. Filtro colaborativo baseado no produto. FunkSVD. Seleção de modelos.

Objetivos:

Fornecer ao aluno a teoria de cada um dos principais métodos de machine learning no enfoque estatístico, bem como mostrar as aplicações práticas nos diversos ramos da ciência destas técnicas. O aluno deverá ser capaz de utilizar os programas de computador relacionados com as técnicas apresentadas.

Metodologia:

Aulas expositivas sobre o conteúdo programático e textos para estudo.

Atividades avaliativas:

O atendimento aos objetivos da disciplina será avaliado através de atividades avaliativas online. As atividades consistirão em trabalhos práticos aplicados à resolução de problemas contextualizados aos diferentes tópicos do conteúdo programático, entregues em formato digital, através da plataforma Moodle.

- Os pontos serão distribuídos da seguinte forma:

  - 5 (cinco) trabalhos práticos (distribuídos ao longo do curso) - VALOR: 100,0 pontos

Será aprovado o aluno que atingir média final igual ou superior a 60,0 pontos. Para não ser reprovado(a) por infrequência e ter possibilidades de realizar o exame especial (caso necessário), o(a) aluno(a) deverá ter pelo menos 75% de frequência nas atividades da disciplina.

Alunos com média final inferior a 60,0 pontos poderão ser aprovados através do Exame Final versando sobre toda a matéria tratada na disciplina.

Obs.: as atividades inerentes à disciplina "**Métodos Estatísticos para Machine Learning**" são de extrema importância para o desenvolvimento ideal do tema.

Todas as atividades propostas nesse curso valem pontuação. Assim sendo, todas as atividades serão pontuadas, somando 100,0 pontos distribuídos na disciplina.

CRONOGRAMA DAS ATIVIDADES AVALIATIVAS

(Sujeito a alterações de acordo com o andamento da disciplina)

Avaliação	Foco	Disponibilização	Entrega	Pontos
Atividade 1	Generalização, flexibilidade, seleção e regularização	16/09/2026	27/09/2026	20
Atividade 2	Métodos locais, kernels e SVM/SVR	30/09/2026	11/10/2026	20
Atividade 3	Modelagem supervisionada completa	21/10/2026	22/11/2026	20
Atividade 4	Redução de dimensionalidade, agrupamento e regras de associação	23/11/2026	06/12/2026	20
Atividade 5	Sistemas de recomendação	07/12/2026	16/12/2026	20
Total				**100**

Observações

- As atividades serão entregues em formato digital, por meio da plataforma Moodle.
- A Atividade 3 será desenvolvida ao longo de um período maior por possuir caráter integrador.
- O período de 17 a 23 de dezembro de 2026 é reservado aos Exames Especiais.

EXAME ESPECIAL: 21/12/2026

CRONOGRAMA DOS CONTEÚDOS

(Sujeito a alterações de acordo com o andamento da disciplina)

Semana	Datas	Conteúdo
1	24/08 e 26/08	Fundamentos do Aprendizado Estatístico
2	31/08 e 02/09	Modelos Lineares para Regressão e Classificação
3	07/09 e 09/09	Funções de Base, Splines e Séries
4	14/09 e 16/09	Seleção e Regularização
5	21/09 e 23/09	Métodos Locais
6	28/09 e 30/09	RKHS, SVR e SVM
7	05/10 e 07/10	Árvores
8	12/10 e 14/10	Bagging e Random Forest
9	19/10 e 21/10	Boosting
10	26/10 e 28/10	Redes Neurais Artificiais
11	02/11 e 04/11	Alta Dimensionalidade
12	09/11 e 11/11	Decisão, Avaliação e Interpretabilidade
13	16/11 e 18/11	Redução de Dimensionalidade
14	23/11 e 25/11	Análise de Agrupamento
15	30/11 e 02/12	Regras de Associação
16	07/12 e 09/12	Sistemas de Recomendação: Filtragem Colaborativa
17	14/12 e 16/12	Sistemas de Recomendação: FunkSVD, Avaliação e Seleção de Modelos

Bibliografia básica

1. Izbicki, R. Santos, T.M., Aprendizado de máquina: uma abordagem estatística. 1ª edição. São Carlos, SP. 2020.
2. James, G., Witten, D., Hastie, T. e Tibshirani. An Introduction to Statistical Learning, 2013
3. Hastie, T., Tibshirani, R. e Friedman, J. The Elements of Statistical Learning, 2009;
4. Lantz, B. Machine Learning with R, Packt Publishing, 2013;
5. Tan, Steinbach, and Kumar. Introduction to Data Mining, Addison-Wesley, 2005.

Bibliografia complementar

1. Chris Bishop (2006). Pattern Recognition and Machine Learning.
2. Luc Devroye, Laszlo Gyorfi, Gabor Lugosi. (1996). A probabilistic theory of pattern recognition.
3. Gyorfi, Kohler, Krzyzak and Walk (2002). A Distribution-Free Theory of Nonparametric Regression.
4. Larry Wasserman (2004). All of Statistics: A Concise Course in Statistical Inference.
5. Larry Wasserman (2005). All of Nonparametric Statistics.