



Universidad Tecnológica de la Mixteca

Clave DGP: 200089

Doctorado en Modelación Matemática

00061

PROGRAMA DE ESTUDIOS

NOMBRE DE LA ASIGNATURA

Modelación para ingeniería de variables

SEMESTRE	CLAVE DE LA ASIGNATURA	TOTAL DE HORAS
Optativa	292918	80

OBJETIVO(S) GENERAL(ES) DE LA ASIGNATURA

Estudiar, analizar y comprender los conceptos básicos para el diseño de modelos en selección y extracción de variables. Con la finalidad de que el estudiante sea capaz de modificar o diseñar nuevos métodos para la resolución de problemas abiertos que se presentan en ciencia de datos.

TEMAS Y SUBTEMAS

1. Selección de variables supervisada

- 1.1. Problemas en selección de características
- 1.2. Ganancia de información IG (Univariado)
- 1.3. Índice GINI
- 1.4. Método de Fisher-Score
- 1.5. CFS (Correlation feature selection multivariado)
- 1.6. Modelos RELIEF (Univariado)
- 1.7. Laplacian Score
- 1.8. Chi-cuadrado
- 1.9. mRMR (mínima redundancia máxima relevancia)

2. Selección de variables no supervisada

- 2.1. Modelos filter, wrapper e híbridos.
- 2.2. Selección espectral de características (SPEC)
- 2.3. Laplacian score (LS)
- 2.4. Selección multicluster (MCFS), NDFS y UDFS
- 2.7. Weighted K-means
- 2.8. Term frequency IDF
- 2.9. Modelo HPStream
- 2.11. Estimación de la máxima verosimilitud.
- 2.12. Redes de Markov y bayesianas.
- 2.13. Modelos basados en vecindades cercanas
- 2.14. Modelos de kernel.

3. Extracción de variables (reducción de dimensionalidad)

- 3.1. Análisis de componentes principales (PCA)
- 3.2. Dual PCA
- 3.3. Kernel PCA
- 3.4. Análisis de correlación canónica (CCA)
- 3.5. Escalada multidimensional (MDS)
- 3.6. ISOMAP
- 3.7. Proyecciones aleatorias
- 3.8. Embebimientos localmente lineales
- 3.9. Clustering espectral
- 3.10. Laplacian eigenmap
- 3.11. Desarrollo de la varianza máxima
- 3.12. t-SNE

ACTIVIDADES DE APRENDIZAJE

Sesiones dirigidas por parte del profesor en la que se presentan los conceptos poniendo énfasis en los fundamentos matemáticos de cada modelo. Se sugiere utilizar algún Notebook como Collaboratory o Jupyter o Visual Studio Code para realizar programas



Universidad Tecnológica de la Mixteca

Clave DGP: 200089

Doctorado en Modelación Matemática

00062

PROGRAMA DE ESTUDIOS

con el lenguaje Python, así como Kaggle y GitHub para compartir y descargar algoritmos programables. Se recomienda ampliamente impartir el curso en un laboratorio con equipo de cómputo disponible para cada estudiante. El contenido será abordado a profundidad y realizada la modelación con fines de práctica de cada uno de los modelos en cada unidad.

CRITERIOS Y PROCEDIMIENTOS DE EVALUACIÓN Y ACREDITACIÓN

Se aplican por lo menos tres exámenes parciales cuyo promedio equivale al 50% de la calificación final, el 50% restante se obtiene de un examen final. Otras actividades que se consideran para la evaluación son las participaciones en clase, asistencias a clases y el cumplimiento de tareas.

BIBLIOGRAFÍA

Básica:

1. B. K. Tripathy, Anveshrithaa Sundareswaran y Shruti Ghela, Unsupervised Learning Approaches for Dimensionality Reduction and Data Visualization, Chapman & Hall/CRC.
2. R. Nilsson, Statistical Feature Selection With Applications in Life Science, Linköping Studies in Science and Technology. Dissertations No. 1090.
3. V. Kumar, Computational Methods of Feature Selection, Data Mining and Knowledge Discovery Series, Chapman & Hall/CRC.
4. M. Kuhn y K. Johnson. Feature Engineering and Selection, A Practical Approach for Predictive Models, CRC Press Taylor & Francis Group.
5. Saúl Solorio-Fernández, J. Ariel Carrasco-Ochoa, y José Fco. Martínez-Trinidad, Hybrid Feature Selection Method for Supervised Classification Based on Laplacian Score Ranking, MCPR 2010, LNCS 6256, pp. 260–269, 2010.

Consulta:

1. I. Guyon y A. Elisseeff, An Introduction to Variable and Feature Selection, Journal of Machine Learning Research 3 (2003) 1157-1182.
2. Liang Sun, Shuiwang Ji, y Jieping Ye. Multi-Label Dimensionality Reduction, Chapman & Hall/CRC.
3. C. C. Aggarwal, Data Classification, Algorithms and Applications, Chapman & Hall/CRC.
4. C. C. Aggarwal, Data Clustering, Algorithms and Applications, Chapman & Hall/CRC.
5. O. Kramer, Dimensionality Reduction with Unsupervised Nearest Neighbors, Springer.

PERFIL PROFESIONAL DEL DOCENTE

Estudios de Doctorado en Matemáticas o en Estadística.

Vo.Bo

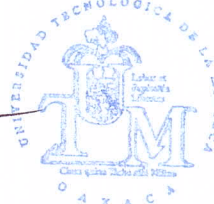
DR. JOSÉ ANIBAL ARIAS AGUILAR
JEFE DE LA DIVISIÓN DE ESTUDIOS
DE POSGRADO



DIVISION DE ESTUDIOS
DE POSGRADO

AUTORIZÓ

DR. RAFAEL MARTÍNEZ MARTÍNEZ
VICE-RECTOR ACADÉMICO



VICE-RECTORIA
ACADÉMICA