

Ciencia de Datos en Ingeniería Civil

1 Descripción del curso

El curso busca introducir a los graduados en ingeniería civil en los conceptos básicos de la ciencia de datos e inferencia estadística, con el objetivo principal de construir, evaluar, y comparar modelos predictivos de fenómenos físicos basados en datos. El alumno será capacitado en la generación de modelos estadísticos que explican los patrones de correlación observados en los datos, así como también en la calibración de modelos computacionales basados en modelos físicos (analíticos y numéricos)

2 Organización

2.1 Docentes

- MSc. Mariano Balbi (mabalbi@fi.uba.ar) ([ResearchGate](#))
- MSc. Santiago Bertero
- Ing. Llamell Martínez Gorbik

2.2 Carga horaria

- Clases: 16 clases de 3 horas en modalidad presencial y virtual sincrónica.
- Trabajos prácticos: 5 Trabajos Prácticos realizados en lenguaje Python, a través de la plataforma online colaborativa Google Colab.

2.3 Evaluaciones

La materia constará de 5 Trabajos Prácticos (TP) y un trabajo final que deberán ser entregados en tiempo y forma para su aprobación.

3 Programa analítico

3.1 Introducción a la probabilidad y estadística

Definiciones de probabilidad. Variables aleatorias, cálculo de probabilidades, probabilidad condicional y teorema de Bayes. Funciones de distribución comunes, y distribuciones multivariadas. Introducción a la estadística bayesiana. Estimación de distribuciones poblacionales, estimadores de media, dispersión y correlación. Intervalos de confianza. Nociones básicas de diseño de experimentos.

3.2 Adquisición y procesamiento de datos

Métodos de obtención de datos típicos en ingeniería civil, y fuentes de datos open-source útiles. Limpieza y homogeneización de datasets: cambios de escala, datos faltantes y outliers. Visualización de datos continuos, discretos y cualitativos en Python: histogramas, boxplots, y gráficos de correlación.

3.3 Calibración de modelos

Definición de modelos y usos en la ingeniería civil. Regresión lineal, cuadrados mínimos y máxima verosimilitud. Inclusión de predictores cualitativos. Regresión multivariada, tratamiento de la multicolinealidad. Extensión a regresión en curvas con polinomios. Regresión no lineal, regresión de modelos físico-matemáticos y modelos computacionales (black-box). Introducción a la regresión de variables discretas: regresión logística. Introducción a modelos no paramétricos: nearest neighbours, procesos gaussianos, y redes neuronales. Algoritmos sencillos de optimización, implementación y uso de paquetes de Python.

3.4 Evaluación y selección de modelos

Diagnóstico de modelos de regresión lineal: evaluación del ajuste y corrección de problemas típicos (heteroscedasticidad, colinealidad de predictores). Evaluación de la precisión de modelos: error de entrenamiento vs error de testeo. Information criteria y validación cruzada para cálculo del error de predicción. Selección de modelos: overfitting y underfitting, y balance entre sesgo y varianza. Optimización de modelos: regularización y selección de variables.

3.5 Cuantificación y propagación de incertidumbres

Descripción y análisis de la incertidumbre en modelos físicos y modelos matemáticos: incertidumbres aleatorias y epistémicas. Análisis de la incertidumbre en los parámetros de los modelos: análisis bayesiano, simulación por cadenas de Markov (MCMC), y método simplificado GLUE para modelos complejos. Propagación de incertidumbre en los parámetros del modelo, a la incertidumbre en las predicciones del modelo: métodos de primer orden (FORM) y simulación de MonteCarlo en modelos computacionales. Introducción a metamodelos (modelos surrogados): interpolación con procesos gaussianos.

3.6 Análisis de sensibilidad

El rol del análisis de sensibilidad en la construcción de modelos. Análisis de sensibilidad local: factor prioritization, factor fixing, método de Morris. Análisis de sensibilidad global: índices de Sobol, índice de efectos totales, filtro de Monte Carlo.

3.7 Riesgo y confiabilidad en ingeniería

Noción de riesgo y confiabilidad en modelos de ingeniería. Modelación de cargas y acciones sobre los modelos. Introducción a procesos de Poisson, homogéneos, y procesos modulados. Noción de tasa de ocurrencia y período de retorno. Obtención de curvas de recurrencia a partir de datos históricos: ajuste de distribución de máximos anuales versus ajuste de modelos Peaks-over-Threshold. Estimación de probabilidad de falla de modelos.

4 Material de estudio

4.1 Material de clase

El contenido del curso está organizado en clases audiovisuales en formato Powerpoint, las que serán entregadas a los alumnos por anticipado.

4.2 Bibliografía

- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). Bayesian Data Analysis, Third Edition. CRC Press.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Springer New York. <https://doi.org/10.1007/978-0-387-84858-7>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning (Vol. 103). Springer New York. <https://doi.org/10.1007/978-1-4614-7138-7>
- McElreath, R. (2020). Statistical rethinking: A Bayesian course with examples in R and Stan (2nd ed.). Taylor and Francis, CRC Press.
- Saltelli, A. (Ed.). (2008). Global sensitivity analysis: The primer. John WileySoftware

5 Software

Se utilizará el lenguaje de programación Python.