

Universidad Nacional Abierta y a Distancia
Vicerrectoría Académica y de Investigación
Curso: Machine Learning
Código: 203008067

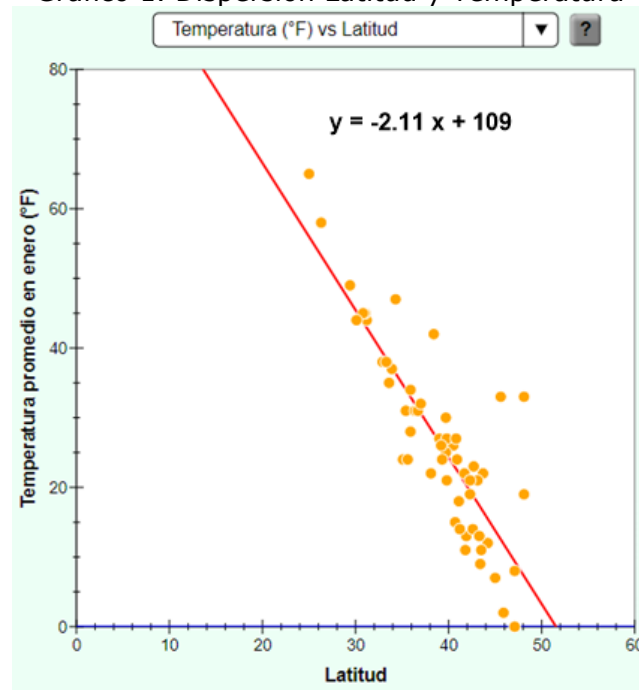
Guía de actividades y rúbrica de evaluación – Fase 2
Aprendizaje Supervisado.

1. Descripción de la actividad

Tipo de actividad: Independiente	
Momento de la evaluación: Intermedio	
Puntaje máximo de la actividad: 50 puntos	
La actividad inicia el: jueves, 22 de febrero de 2024	La actividad finaliza el: jueves, 21 de marzo de 2024
Con esta actividad se espera conseguir los siguientes resultados de aprendizaje: Examinar los conceptos y modelos del aprendizaje automático para aplicarlos en la disciplina de desempeño.	
La actividad consiste en: Desarrollar la siguiente serie de actividades (No es obligatorio programar o el uso de softwares) .	
Actividad 1: Regresión Lineal y Logística.	
a) Realice un glosario con los siguientes conceptos: Variable dependiente, Variable independiente, Descenso del gradiente, Función de pérdida.	
b) Responda a las siguientes preguntas, basados en el gráfico 1.	
<ul style="list-style-type: none">Identifique la variable dependiente y la variable independiente. ¿Cuál es el propósito de la regresión lineal en este caso específico?¿Cuál es el modelo o representación matemática que se obtuvo?	

- Interprete la pendiente.
- Mencione dos métodos para calcular las estimaciones de los parámetros del modelo.
- ¿Cuál es la diferencia entre un modelo lineal simple y un modelo lineal múltiple?
- ¿Cuál es la función de pérdida que se intenta minimizar? Expréselo en sus palabras (ayuda: ECM).

Gráfico 1. Dispersión Latitud y Temperatura

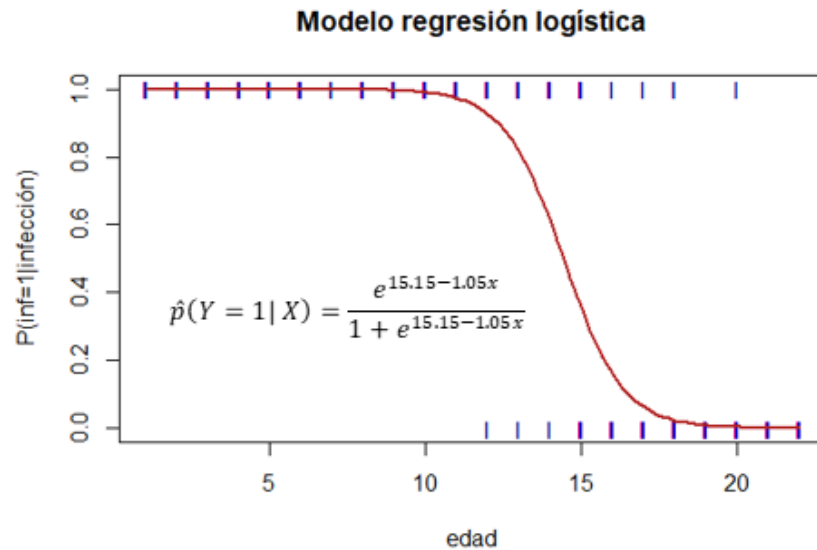


Nota. Fuente: https://phet.colorado.edu/sims/html/least-squares-regression/latest/least-squares-regression_all.html?locale=es

c) Use el gráfico 2 para responder a las siguientes preguntas:

- ¿Por qué un modelo de regresión lineal simple no es una herramienta apropiada para describir la relación entre edad del paciente y la presencia de infección?
- ¿Cuándo se debe utilizar regresión lineal y cuando regresión logística?
- Describa un objetivo del análisis de regresión logística en este caso en particular.

Gráfico 2. Infecciones asociadas a dispositivos notificados en SIVIGIL



Nota: Elaborado a partir de datos Fuente MinTIC.
<https://www.datos.gov.co/dataset/Infecciones-asociadas-a-dispositivos-notificados-e/wmtg-fef3>

Actividad 2: Descenso máximo del gradiente.

a) Responda:

¿Qué es el gradiente de una función?

¿Recuerda alguna aplicación del gradiente?

b) Suponga que se tiene una función de costo dada por la siguiente expresión $f(x) = 3x^2 - 7x + 1$.

- Encuentre el mínimo de la función por el método analítico de las derivadas.
- Encuentre una aproximación del mínimo usando el método máximo del gradiente. Para ello suponga una tasa de aprendizaje $\alpha = 0.1$ y valor inicial $x^0 = 0$. Calcule las primeras seis iteraciones.

Fórmula recurrente del algoritmo $x^{k+1} = x^k - \alpha g(x^k)$
 $g(x)$ es el gradiente de $f(x)$.

Actividad 3: Árboles de decisión.

- Realice un glosario con los siguientes conceptos:
Nodo, Arcos, Hojas, Ramas.
- Con base en los datos de la tabla 1, diseñe un árbol de decisión (manualmente) para clasificar si se realiza o no una actividad al aire libre. La clasificación tiene en cuenta criterios como el nivel de energía requerido, el tamaño de los grupos y si se requiere equipo.

Tabla 1. Respuestas de individuos respecto a una actividad al aire libre

Ind	Energía	Grupo	Equipo	Actividad al aire libre
I1	Mucha	Grande	No	No
I2	Mucha	Grande	Si	No
I3	Poca	Grande	No	Si
I4	Poca	Grande	No	Si
I5	Poca	Pequeño	No	Si
I6	Poca	Pequeño	Si	No
I7	Poca	Pequeño	Si	Si
I8	Mucha	Grande	No	No
I9	Mucha	Pequeño	No	Si
I10	Poca	Pequeño	No	Si
I11	Mucha	Pequeño	Si	Si
I12	Poca	Grande	Si	Si
I13	Poca	Pequeño	No	Si
I14	Poca	Grande	Si	No

Nota: Fuente propia.

- ¿Este árbol es de clasificación o de regresión, explique?
- ¿Con los mismos datos se pueden realizar dos árboles de decisión diferentes? ¿Cómo se decide cuál es el primer nodo?

Actividad 4: Método del vecino más cercano (KNN).

a) Realice un glosario con los siguientes conceptos:

¿Cómo se determina un vecino más cercano en un dataset?

¿Qué métricas existen si las clases son numéricas?

¿Qué métricas existen si las clases son categóricas?

b) Describa un problema típico de su disciplina de desempeño profesional donde:

- Deba tomar una decisión a partir de información dada por dos categorías.
- Indique dos características de tipo cuantitativo.
- Describa cómo sería el uso del algoritmo de KNN en este escenario, para clasificar un nuevo registro, si se usa la distancia euclidiana y se elige $k=3$ como el número de vecinos más cercanos.

Para el desarrollo de la actividad tenga en cuenta que:

En el entorno de Información inicial debe:

- Revisar la Agenda del curso.
- Revisar las subsecciones de Acompañamiento docente.

En el entorno de Aprendizaje debe:

- Revisar las referencias de la Unidad 1.
- Interactuar, en el foro de la Fase 2, con su tutor y sus compañeros de grupo sobre las temáticas abordadas.
- Presentar, en el foro de la Fase 2, los avances de cada una de las actividades.

En el entorno de Evaluación debe: Entregar un documento en PDF nombrado Fase_2_(nombre del estudiante) con las siguientes características:

- Portada.
- Resumen.
- Desarrollo de las cuatro actividades.
- Análisis de resultados de las cuatro actividades.

- Referencias bibliográficas en normas APA (versión 7).

Evidencias de trabajo independiente: Las evidencias de trabajo independiente para entregar son: Entrega de trabajo escrito consolidado en PDF con el desarrollo de las cuatro actividades.

Evidencias de trabajo grupal: En esta actividad no se requieren evidencias de trabajo grupal.

2. Lineamientos generales para la elaboración de las evidencias de aprendizaje a entregar.

Para evidencias elaboradas **independientemente**, tenga en cuenta las siguientes orientaciones

1. Todos los integrantes del grupo deben participar con sus aportes en el desarrollo de la actividad.
2. Identifique los recursos y referentes de la unidad a la que corresponde la fase.
3. Intervenga en el foro de discusión aplicando las normas de netiqueta virtual, evidenciando siempre respeto por las ideas de sus compañeros y del cuerpo docente.
4. Antes de entregar el producto solicitado deben revisar que cumpla con todos los requerimientos que se señalaron en esta guía de actividades, rúbrica de evaluación y por parte del tutor en el foro de discusión.
5. No cometa fraudes, ni plagios ni actos que atenten contra el normal desarrollo académico de las actividades.

Tenga en cuenta que todos los productos escritos individuales o grupales deben cumplir con las normas de ortografía y con las condiciones de presentación que se hayan definido.

En cuanto al uso de referencias considere que el producto de esta actividad debe cumplir con las normas **APA**

En cualquier caso, cumpla con las normas de referenciación y evite el plagio académico, para ello puede apoyarse revisando sus productos escritos mediante la herramienta Turnitin que encuentra en el campus virtual.

Considere que en el acuerdo 029 del 13 de diciembre de 2013, artículo 99, se considera como faltas que atentan contra el orden académico, entre otras, las siguientes: literal e) "El plagiar, es decir, presentar como de su propia autoría la totalidad o parte de una obra, trabajo,

documento o invención realizado por otra persona. Implica también el uso de citas o referencias faltas, o proponer citad donde no haya coincidencia entre ella y la referencia” y liberal f) “El reproducir, o copiar con fines de lucro, materiales educativos o resultados de productos de investigación, que cuentan con derechos intelectuales reservados para la Universidad”

Las sanciones académicas a las que se enfrentará el estudiante son las siguientes:

- a) En los casos de fraude académico demostrado en el trabajo académico o evaluación respectiva, la calificación que se impondrá será de cero puntos sin perjuicio de la sanción disciplinaria correspondiente.
- b) En los casos relacionados con plagio demostrado en el trabajo académico cualquiera sea su naturaleza, la calificación que se impondrá será de cero puntos, sin perjuicio de la sanción disciplinaria correspondiente.

3. Formato de Rúbrica de evaluación

Tipo de actividad: Independiente	
Momento de la evaluación: Intermedio	
La máxima puntuación posible es de 50 puntos	
<p>Primer criterio de evaluación:</p> <p>Interpreta Modelos lineales usando regresión lineal y logística</p> <p>Este criterio representa puntos del 12 total de 50 puntos de la actividad.</p>	<p>Nivel alto: El estudiante interpreta todos los diferentes modelos lineales solicitados ajustados a los datos proporcionados. Comparte de manera oportuna sus avances en el foro colaborativo. Si su trabajo se encuentra en este nivel puede obtener entre 10 puntos y 12 puntos</p> <p>Nivel Medio: El estudiante no interpreta los dos modelos lineales ajustados a los datos proporcionados y/o solo interpreta un modelo lineal y/o no realiza los ajustes sugeridos por el tutor y/o no comparte de manera oportuna sus avances en el foro colaborativo Si su trabajo se encuentra en este nivel puede obtener entre 7 puntos y 11 puntos</p> <p>Nivel bajo: El estudiante no interpreta los dos modelos lineales y no participa de manera oportuna en el foro colaborativo. Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 6 puntos</p>
<p>Segundo criterio de evaluación:</p> <p>Implementa el método del descenso máximo del gradiente</p> <p>Este criterio representa 10 puntos del total de 50 puntos de la actividad</p>	<p>Nivel alto: El estudiante implementa método del descenso máximo del gradiente, incluye el análisis sustentando por los resultados y la gráfica del método. Comparte de manera oportuna sus avances en el foro colaborativo. Si su trabajo se encuentra en este nivel puede obtener entre 9 puntos y 10 puntos</p> <p>Nivel Medio: El estudiante implementa método del descenso máximo del gradiente, sin embargo, el análisis no está sustentando por los resultados y/o no incluye la gráfica del método y/o no realiza los ajustes sugeridos por el tutor y/o no comparte de manera oportuna sus avances en el foro colaborativo Si su trabajo se encuentra en este nivel puede obtener entre 6 puntos y 8 puntos</p>

	<p>Nivel bajo: El estudiante no implementa método del descenso máximo del gradiente y no participa de manera oportuna en el foro colaborativo.</p> <p>Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 5 puntos</p>
<p>Tercer criterio de evaluación:</p> <p>Aplica árboles de decisión.</p> <p>Este criterio representa 10 puntos del total de 50 puntos de la actividad</p>	<p>Nivel alto: El estudiante aplica árboles de decisión a cada uno de los modelos propuestos. Comparte de manera oportuna sus avances en el foro colaborativo.</p> <p>Si su trabajo se encuentra en este nivel puede obtener entre 9 puntos y 10 puntos</p> <p>Nivel Medio: El estudiante aplica árboles de decisión, sin embargo, no incluye una gráfica que modele el problema y/o no realiza los ajustes sugeridos por el tutor y/o no comparte de manera oportuna sus avances en el foro colaborativo.</p> <p>Si su trabajo se encuentra en este nivel puede obtener entre 6 puntos y 8 puntos</p> <p>Nivel bajo: El estudiante no aplica árboles de decisión a modelos clasificación y no participa de manera oportuna en el foro colaborativo.</p> <p>Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 5 puntos</p>
<p>Cuarto criterio de evaluación:</p> <p>Aplica el algoritmo KNN.</p> <p>Este criterio representa 10 puntos del total de 50 puntos de la actividad</p>	<p>Nivel alto: El estudiante describe un problema de su disciplina profesional y aplica correctamente el algoritmo k-NN. Comparte de manera oportuna sus avances en el foro colaborativo.</p> <p>Si su trabajo se encuentra en este nivel puede obtener entre 9 puntos y 10 puntos</p> <p>Nivel Medio: El estudiante describe un problema de su disciplina profesional pero no aplica correctamente el algoritmo KNN y/o no realiza los ajustes sugeridos por el tutor y/o no comparte de manera oportuna sus avances en el foro colaborativo.</p>

	<p>Si su trabajo se encuentra en este nivel puede obtener entre 6 puntos y 8 puntos</p> <p>Nivel bajo: El estudiante no describe un problema de su disciplina profesional ni aplica el algoritmo KNN y no participa de manera oportuna en el foro colaborativo.</p> <p>Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 5 puntos</p>
<p>Quinto criterio de evaluación:</p> <p>Documento solicitado en el entorno de Evaluación.</p> <p>Este criterio representa 8 puntos del total de 50 puntos de la actividad</p>	<p>Nivel alto: El estudiante entrega el documento en el entorno de evaluación, todos los elementos solicitados.</p> <p>Si su trabajo se encuentra en este nivel puede obtener entre 7 puntos y 8 puntos</p> <p>Nivel Medio: Entrega el documento en el entorno de evaluación, pero no cumple con todos los elementos solicitados.</p> <p>Si su trabajo se encuentra en este nivel puede obtener entre 4 puntos y 6 puntos</p> <p>Nivel bajo: No presenta el documento consolidado en el entorno de Evaluación con todos los elementos solicitados.</p> <p>Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 3 puntos</p>