

## TABLA DE CONTENIDOS

	página
<b>Dedicatoria</b>	<b>I</b>
<b>Agradecimientos</b>	<b>II</b>
<b>Tabla de Contenidos</b>	<b>III</b>
<b>Índice de Figuras</b>	<b>VI</b>
<b>Índice de Tablas</b>	<b>VIII</b>
<b>Resumen</b>	<b>IX</b>
<b>Abstract</b>	<b>X</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Descripción de la Propuesta . . . . .	1
1.1.1. Contexto del Proyecto . . . . .	2
1.1.2. Trabajo Relacionado . . . . .	3
1.1.3. Definición del Problema . . . . .	3
1.1.4. Propuesta de Solución . . . . .	4
1.2. Hipótesis . . . . .	6
1.3. Objetivos . . . . .	6
1.4. Alcances . . . . .	7
<b>2. Marco Teórico</b>	<b>8</b>
2.1. Mejoramiento Genético del Trigo . . . . .	8
2.2. Reflectancia Espectral . . . . .	9
2.3. Análisis de Regresión . . . . .	11
2.3.1. ¿Qué es regresión? . . . . .	12
2.3.2. Multiple Linear Regression . . . . .	13
2.3.3. Partial Least Squares Regression . . . . .	14
2.3.4. Ridge Regression . . . . .	15
2.3.5. Support Vector Regression . . . . .	16

2.4.	Otros Algoritmos Utilizados . . . . .	19
2.4.1.	Principal Component Analysis (PCA) . . . . .	19
2.4.2.	Local Outlier Factor (LOF) . . . . .	21
2.5.	Medidas de Rendimiento . . . . .	22
<b>3.</b>	<b>Descripción de la Librería (API)</b>	<b>25</b>
3.1.	Subproblemas en Minería de Datos . . . . .	25
3.1.1.	Valores Faltantes . . . . .	25
3.1.2.	Normalización . . . . .	26
3.2.	Detección de Ruido . . . . .	27
3.3.	Gráficos de Diagnóstico . . . . .	31
3.4.	Otros Gráficos . . . . .	34
3.4.1.	Gráfico del Codo: Elbow Plot . . . . .	34
3.4.2.	Gráfico de Dispersión . . . . .	35
3.4.3.	Gráfico de Dispersión Matricial . . . . .	37
3.4.4.	Gráfico 3D . . . . .	38
3.4.5.	Gráfico de Densidad . . . . .	39
3.4.6.	Ridge Plot . . . . .	41
<b>4.</b>	<b>Metodología Experimental</b>	<b>44</b>
4.1.	Adquisición de Datos . . . . .	45
4.2.	Variables de Estudio . . . . .	46
4.3.	Pre-procesamiento . . . . .	47
4.3.1.	Datos Faltantes . . . . .	48
4.3.2.	Detección de Ruido . . . . .	48
4.3.3.	Detección de Datos Atípicos . . . . .	50
4.3.4.	Normalización de Datos . . . . .	52
4.3.5.	Reducción de la Dimensionalidad . . . . .	53
4.4.	Modelado Predictivo . . . . .	53
4.5.	Validación del Modelo . . . . .	56
<b>5.</b>	<b>Resultados</b>	<b>58</b>
<b>6.</b>	<b>Conclusiones</b>	<b>68</b>

<b>Bibliografía</b>	<b>71</b>
<b>Anexos</b>	
<b>A: Resultados 2011 y 2012</b>	<b>76</b>

## ÍNDICE DE FIGURAS

	página
2.1. Ejemplo de reflectancia en una hoja . . . . .	10
2.2. Rangos del espectro y características medibles en las plantas . . . . .	11
2.3. Plano que se ajusta a los datos cuando $\mathcal{X} \in \mathbf{R}^2$ . . . . .	13
2.4. Hiperplano que clasifica correctamente y maximiza el margen . . . . .	16
2.5. Ejemplo de una transformación del espacio de entrada . . . . .	17
2.6. SVR que muestra la holgura de la función de pérdida $\epsilon$ . . . . .	19
2.7. Componentes principales para un distribución de ejemplo . . . . .	20
2.8. Componentes principales para un distribución de ejemplo . . . . .	21
3.1. Ejemplo de la firma espectral de un subconjunto de los datos . . . . .	28
3.2. Gráficos de influencia, leverage y residuos para el conjunto de datos <i>iris</i> . . . . .	32
3.3. Gráfico del codo de los componentes principales del conjunto de datos <i>cars</i> . . . . .	35
3.4. Gráfico de dispersión de los 2 primeros componentes principales del conjunto de datos <i>cars</i> . . . . .	36
3.5. Gráfico de dispersión matricial de los primeros 4 componentes princi- pales del conjunto de datos <i>cars</i> . . . . .	38
3.6. Gráfico de densidad para los puntajes resultantes de LOF para el conjunto de datos <i>cars</i> . . . . .	40
3.7. Gráfico de ejemplo de optimización del parámetro $\lambda$ . . . . .	42
4.1. Arquitectura del modelado predictivo . . . . .	45
4.2. Zonas comunes con ruido en los datos de reflectancia espectral . . . . .	49
4.3. Gráfico de densidad con las puntuaciones asignadas por el algoritmo LOF . . . . .	52
4.4. Gráfico del codo de los componentes principales de un conjunto de datos	54
4.5. Gráfico de optimización del parámetro $\lambda$ . . . . .	55
4.6. Ejemplificación del funcionamiento de 10 veces validación cruzada . . . . .	56
5.1. Datos predichos vs observados después de la validación cruzada para rendimiento . . . . .	65

5.2. Datos predichos vs observados después de la validación cruzada para IAF . . . . .	66
5.3. Datos predichos vs observados después de la validación cruzada para $\Delta^{13}C$ . . . . .	67

## ÍNDICE DE TABLAS

	página
4.1. Conjuntos de datos de estudio . . . . .	46
5.1. Media y desviación estándar del estadístico $R^2$ para los cuatro modelos con datos del 2011 . . . . .	59
5.2. Media y desviación estándar del estadístico $R^2$ para los cuatro modelos con datos del 2012 . . . . .	60
5.3. Detalle de los resultados obtenidos para los 3 rasgos con mejores ren- dimiento durante el 2011 . . . . .	62
5.4. Detalle de los resultados obtenidos para los 3 rasgos con mejores ren- dimiento durante el 2012 . . . . .	63
A.1. Resultados estadísticos de la predicción de los rasgos evaluados en los 4 ambientes, para los 2 estados fenológicos medidos durante el 2011 para los 4 modelos de regresión evaluados . . . . .	76
A.2. Resultados estadísticos de la predicción de los rasgos evaluados en los 4 ambientes, para los 2 estados fenológicos medidos durante el 2012 para los 4 modelos de regresión evaluados . . . . .	81