

# Conclusiones

*“Para surcar mejores aguas despliega ahora las velas la navecilla  
de mi ingenio, que deja tras de sí un mar tan cruel”*

Dante Alighieri.

El estudio realizado sobre la definición del problema general de aprendizaje a partir de ejemplos descrito como la minimización de un funcional de riesgo sobre la base de datos empírico permite concluir que

- Según sea la naturaleza de los valores en los datos de salida, el problema general da lugar al enunciado de diferentes tareas de aprendizaje. En el estudio se han particularizado los casos más estándares o tradicionales como la estimación de regresiones, el reconocimiento de patrones o la regresión ordinal, si bien existen otras aproximaciones que podrían ser consideradas como la detección ‘novedosa’ — del inglés, *novelty detection* —, el aprendizaje sobre datos no etiquetados o sobre información negativa.
- La elección del principio inductivo sobre el que se sienta la base del método de aprendizaje resultará crítica para conseguir el resultado deseado, en el caso de la presente tesis la propiedad de generalización sobre nuevas entradas. A tal efecto, se ha mostrado como el principio de regularización o penalización y el principio de minimización del riesgo estructural se perfilan desde un punto de vista teórico como aquellos más adecuados en la obtención de soluciones generalizadoras.
- Por último, debe ser elegido el método de aprendizaje que llevará a la práctica el principio inductivo. Al respecto, han sido enunciadas algunas metodologías generales referentes a la técnica de regularización y a la aplicación de la minimización del riesgo estructural.

Debe reflexionarse en este momento sobre la elección del método de aprendizaje como una consecuencia natural de unos principios teóricos bien fundamentados respecto a los fines que se persiguen. No se pretende justificar la elección *a posteriori*

en función de unos resultados aceptables sobre una colección de ‘benchmarks’ o problemas tipo, resultados que también se consiguen, puesto que en la mayoría de casos éstos dependen más de la habilidad del investigador en la selección de parámetros que en la capacidad de la máquina de aprendizaje.

De los principios inductivos señalados anteriormente como más adecuados, la regularización es una técnica ampliamente aplicada para trabajar con problemas mal situados y mal condicionados. Del estudio realizado sobre este principio se concluye que

- Su análisis desde un doble enfoque, la visión bayesiana y la versión como proceso variacional, ha significado una nueva aportación que permite demostrar que si son utilizados como elementos de regularización funcionales definidos en teoría de ‘splines’ es posible reducir el cálculo de la matriz de regularización a un proceso directo y exacto, evitando los procedimientos iterativos que hasta el momento se hacían necesarios y que resultaban tan caros desde el punto de vista computacional. En particular se ha demostrado en esta memoria que la regularización de segundo orden puede ser obtenida desde un marco de trabajo variacional realizando aproximaciones más o menos precisas para calcular la matriz Hessiana de derivadas de orden dos por lo que la precisión final se ve mermada. El nuevo planteamiento ofrece una fácil y eficiente expresión para el cálculo de esta matriz.
- La comunidad investigadora ha realizado un mayor esfuerzo investigador dirigido al cálculo sobre la intensidad de regularización que debe poseer una solución, sin embargo el hecho de que la función a ser aproximada sea una función desconocida provoca que este trabajo comparativo en busca de la determinación del mejor parámetro para un problema específico posea siempre un alto grado de incertidumbre motivada por el ruido en las observaciones o la ausencia de datos. Como reflexión, apuntar que la voluntad de obtener una precisión cada vez mayor en el proceso de aproximación a una función desconocida ruidosa mediante la determinación de un parámetro exacto es un propósito poco realista.
- La herramienta clave para la transferencia de información entre las visiones bayesiana y variacional es el operador regularizador. Una importante tarea matemática a realizar como extensión de trabajo es la de buscar operadores adecuados a las características deseadas por el operador.

Las redes de regularización, fruto del principio inductivo de penalización, son una herramienta válida de trabajo, pero sus propiedades de aproximación exclusivamente asintóticas y la expansión de la función solución sobre un elevado número de vectores las convierten en poco prácticas. La búsqueda de soluciones ha hecho fijar los ojos de algunos investigadores en el buen comportamiento de las máquinas de soporte vectorial. Este tipo de máquinas poseen además la característica de estar basadas en un

principio inductivo, el SRM, que tiene en cuenta la finitud del conjunto de entrenamiento y su construcción se desarrolla sobre espacios de aproximación anidados.

El estudio de las relaciones entre ambos principios inductivos, en un campo de investigación aún abierto y altamente dinámico, buscando establecer un marco unificado permite concluir que

- Es posible derivar tanto las redes de regularización como las máquinas de soporte vectorial aceptando algunos condicionantes sobre la forma de las funciones de Green.
- El análisis crítico y estructuración realizados en esta memoria de los resultados obtenidos hasta el momento y el punto de vista que brinda el marco SRM parecen ofrecer mayores garantías teóricas en la búsqueda de una solución que generalice bien sobre problemas basados en un conjunto de datos finito.

Cuando se intenta llevar a la práctica las SVMs en problemas de clasificación extraídos de la vida real es común encontrarse con situaciones que desde el punto de vista binario de la clasificación SVM se definen como de multclasificación. En este sentido, se ha realizado una búsqueda exhaustiva y estudio detallado sobre técnicas de reducción de problemas multiclase sobre particiones binarias llegándose a las conclusiones

- El estudio del problema de multclasificación acaba basándose sobre problemas binarios: al utilizar máquinas biclasificadoras se afirma tradicionalmente que su extensión a entornos multiclase es “*evidente*”; cuando la máquina ya está concebida para el trabajo con múltiples salidas el grueso del desarrollo teórico se reduce al caso binario pues la generalización a casos multiclase se hace “*obvia*”; por último, existe un tercer tipo de máquina, por ejemplo los árboles de decisión, que trabajan directamente con problemas multiclase aunque para ello necesiten de unos nodos de decisión que son, en la mayor parte de las ocasiones, dicotomías.
- Se ha demostrado en esta tesis que la arquitectura final de un clasificador global fusión de clasificadores parciales depende de: el acondicionamiento, interpretación y/o agrupación de las entradas multiclase para realizar las clasificaciones binarias; el tipo o tipos de máquina de aprendizaje elegido para realizar la descomposición del espacio de clasificación en biparticiones; el esquema de reconstrucción que realiza la fusión de las predicciones; la interpretación de la respuesta facilitada por el clasificador global. Para tener en cuenta todos estos elementos se ha desarrollado una notación original que permite la comparación de las arquitecturas existentes y facilita la comprensión de su funcionamiento.

- Las arquitecturas multiclase de descomposición y reconstrucción con esquema de descomposición en paralelo son las más utilizadas cuando se trabaja con nodos de dicotomía tipo SVMC. En el método de descomposición, es posible codificar la información en elementos binarios  $\{-1, +1\}$  mediante tres esquemas principalmente: los más tradicionales 1- $v$ - $r$  y 1- $v$ -1, y el ECOC.
- Las SVMs multiclase desarrolladas en los últimos años considerando todas las clases a la vez durante el proceso de aprendizaje necesitan de un coste computacional muy significativo. Para su análisis, se ha realizado el esfuerzo de modificar la notación original de forma que ha sido posible unificar enunciados y ejemplarizar de forma clara sus puntos de similitud y su capacidad de generalización.
- Se ha definido como aportación una nueva herramienta de análisis, un factor de robustez que permite analizar la posibilidad de incorporar fallos en los nodos de dicotomía sin que el resultado correcto final se vea afectado. Aunque la definición no es aplicable en todas las arquitecturas, se ha observado la nula robustez de la mayoría de máquinas, mientras que sobre aquellas consideradas más robustas, las de estructura ECOC, no es posible calcular este factor en forma general debido a la aleatoriedad en la elección de los nodos de dicotomía.

El estudio previo sobre el uso de SVMs en entornos multiclase, en especial en esquemas estándares de descomposición y reconstrucción lleva a la conclusión de que

- Un esquema de descomposición 1- $v$ -1 no permite tener en consideración todas las entradas de entrenamiento y se ve obligada a emitir una predicción aunque no haya sido entrenada sobre la clase a la que pertenece la entrada a validar.
- Un esquema de descomposición 1- $v$ - $r$  no permite concentrar el esfuerzo clasificador.
- Los esquemas de descomposición ECOC permiten aumentar la robustez de la salida pero deben ser considerados en la mayoría de ocasiones de forma aleatoria y su eficacia depende de la asignación de clases.

por lo que se hace necesaria la definición de un nuevo algoritmo que considere desde su base el uso que se hará de él en esquemas de múltiples salidas. A tal efecto se ha desarrollado como principal aportación el denominado algoritmo  $K$ -SVCR. El análisis detallado de su funcionamiento permite afirmar que

- La formulación matemática del problema QP asociado permite interpretar el nuevo algoritmo como un intermedio entre el método SVMC y el método SVMR.
- La definición del nuevo algoritmo permite integrar los esquemas de descomposición 1- $v$ -1 y 1- $v$ - $r$  superando algunas de sus desventajas iniciales.

- El algoritmo es más eficiente que una SVMR con tres salidas gracias a su menor demanda en las restricciones del problema QP asociado.
- El nuevo factor de insensitividad resulta crítico en la definición de la tercera clase, su amplitud y el número de vectores soporte.

La capacidad de aprendizaje de la nueva máquina, aún siendo definida como una entidad independiente, se muestra de forma completa sobre problemas de clasificación multiclase siguiendo un esquema de {descomposición, reconstrucción}. Las conclusiones a las que se llega tras su puesta en práctica son

- La utilización de todas las clases en el entrenamiento de todos los nodos de dicotomía, aún aumentando el tiempo de cómputo, consigue la aparición de una cierta redundancia que le confiere a esta nueva formulación una tolerancia a fallos muy superior a la obtenida con arquitecturas convencionales.
- Para hacer plausible esta robustez se hace necesario definir un elemento intérprete y un elemento de combinación de predicciones que favorezcan esta característica, los cuales han sido definidos y verificadas sus prestaciones en esta memoria.
- El comportamiento robusto de la teoría ECOC sobre máquinas triclase es cualitativamente superior a aquél obtenido sobre máquinas biclase.

Tras analizar el uso de penalizadores en problemas de regresión y crear una nueva máquina de aprendizaje bajo la teoría de aprendizaje estadístico para el caso de clasificación múltiple, se analiza la posibilidad de extender esta última para el caso de regresión ordinal u ordenación. Los estudios realizados en la búsqueda de una solución correcta lleva a las siguientes conclusiones

- Es necesario modificar la función de coste en el sentido de conseguir que la nueva máquina aprenda preferencias.
- Una comparativa con otros resultados existentes basados en esta modificación muestra que la nueva aportación trabaja sobre un conjunto de aprendizaje modificado más reducido.
- En cualquier caso, se produce un significativo incremento de la dimensionalidad del nuevo conjunto de patrones en forma de pares respecto al original.

Las particulares condiciones de trabajo deseables para este tipo de problema son cumplidas por el problema de riesgo crediticio analizado como aplicación y cuyos resultados sirven para mostrar la capacidad generalizadora del nuevo método. La segunda de las aplicaciones, en sistemas de depuración de aguas residuales, también muestra la efectividad de la nueva arquitectura en entornos donde la variable temporal es significativa en el resultado de la identificación.

