# FUNDAMENTOS TEÓRICOS

## SISTEMAS DINÁMICOS

Un sistema es aquel en donde los elementos están dentro de una organización, en la cual estos se encuentran interrelacionados e interactúan entre sí. Existen dos tipos de sistemas, el estático y el dinámico.

Un sistema dinámico, a diferencia del estático, es aquel que considera la historia, su estado evoluciona en el tiempo, es decir, incorpora la dimensión temporal para determinar un ordenamiento espacial de los datos. El orden temporal de los datos presenta información importante, no es solo un conjunto de datos ordenados. En la práctica, esto se traduce en que la salida en el tiempo *t* de un sistema dinámico depende de la entrada u en el mismo tiempo, además de las salidas y entradas en tiempos anteriores. Los datos ordenados a lo largo del tiempo son las denominadas series de tiempo *f(t).*

Sistema

u(t)

y(t)

Ilustración 2‑1 Sistema dinámico

Un modelo dinámico puede ser caracterizado como lineal o no lineal. Dentro de los modelos lineales clásicos utilizados en el estudio del SAC destacan los modelos de *Aslid-Tiecks*, los cuales utilizan solo la PAM como variable de entrada y permite clasificar la calidad de la autorregulación a través del índice de autorregulación, denominado ARI. En 1998 Panerai realizó un modelo lineal para estimar la regulación cerebral dinámica basado en variaciones espontáneas de la PAM, y demostró que pequeñas fluctuaciones en la PIC tienen influencias importantes en la PPC. Dentro de los métodos no lineales algunos modelos que destacan son Wiener Laguerre (Panerai et al., 1999; Panerai et al., 2001), las redes neuronales (muñoz, 2004; Panerai et al., 2004, Chacón et al., 2005), y finalmente algunos trabajos realizados por medio de SVM (Días, 2005; Araya, 2006; Bello, 2007; Ruz, 2009; Muñoz, 2009; Sanhueza 2011; Varas, 2013; Sun-Ho, 214), entre otros.

Una de las mayores dificultades es el desarrollo de una regla que regule todas las mediciones, estándares de medición, *“gold standards”.* Aaslid et al. (1989) desarrolló un índice con el objetivo de medir y cuantificar la eficacia de la autorregulación dinámica del *flujo sanguíneo cerebral* (FSC) que será descrito a continuación.

### Modelo Aaslid-Tiecks (A-T).

Este modelo ha sido propuesto por Aaslid - Tiecks (Tiecks *et al.*, 1995), y fue especialmente diseñado para modelar y clasificar la *velocidad de flujo sanguíneo cerebral* (VFSC). Éste corresponde al modelo de estados de segundo orden que se muestra en las ecuaciones 3.1 a 3.4.

|  |  |  |
| --- | --- | --- |
|  |  | *(3.1)* |
|  |  | *(3.2)* |
|  |  | *(3.3)* |
|  |  | *(3.4)* |
| En donde | *dP* es el cambio normalizado de *PAM* en torno a la presión de control (*P0*), considerando el efecto de la presión de cierre crítica (*PCC*).  La *V0* es la velocidad de control en la arteria cerebral media.  *T* es la constante de tiempo.  *D* es el factor de amortiguamiento  *K* es la ganancia autorregulatoria dinámica.  *f* es la frecuencia de muestreo.  *x1* y *x2* son variables de estado que se asumen iguales a cero durante el periodo de control. | |

A partir de las señales reales de presión y con diez tripletas de valores para los parámetros *D*, *K* y *T*, se generan diez curvas teóricas que representan distintos niveles de comportamiento del SAC expresadas en el ARI, que toma valores ordinales desde 0 (autorregulación absolutamente deteriorada) hasta 9 (autorregulación perfecta).

La Tabla 3.1 muestra los valores que toman los parámetros, la Figura 3.1 muestra las diez respuestas teóricas al escalón inverso de presión. En ella se observan claramente los diez niveles de calidad autorregulatoria.

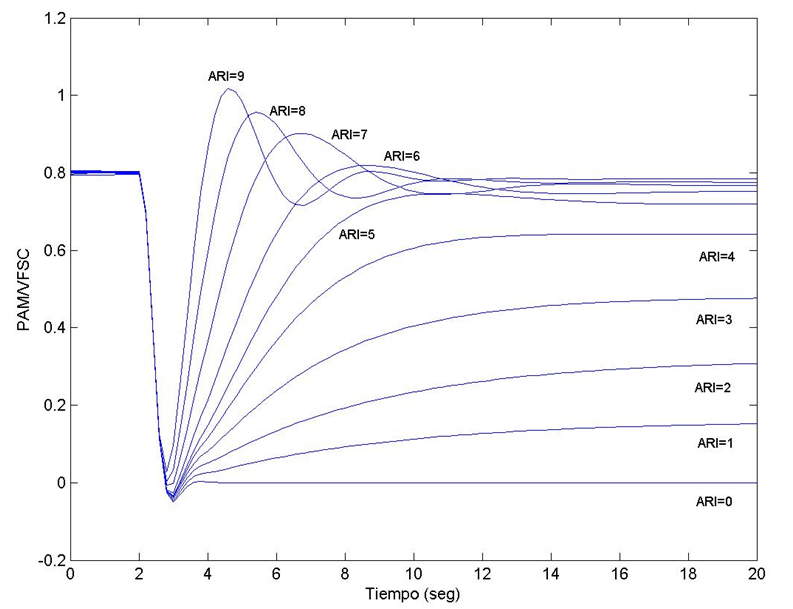
*Tabla 3.1. Parámetros utilizados en el Modelo Aaslid-Tiecks.*

|  |  |  |  |
| --- | --- | --- | --- |
| ***T*** | ***D*** | ***K*** | ***ARI*** |
| *2.00* | *1.70* | *0.00* | *0* |
| *2.00* | *1.60* | *0.20* | *1* |
| *2.00* | *1.50* | *0.40* | *2* |
| *2.00* | *1.15* | *0.60* | *3* |
| *2.00* | *0.90* | *0.80* | *4* |
| *1.90* | *0.75* | *0.90* | *5* |
| *1.60* | *0.65* | *0.94* | *6* |
| *1.20* | *0.55* | *0.96* | *7* |
| *0.87* | *0.52* | *0.97* | *8* |
| *0.65* | *0.50* | *0.98* | *9* |

Cada una de las diez curvas teóricas se compara con la señal de VFSC real, asignando así a esta última el ARI correspondiente a la curva más ajustada (Tiecks *et al.*, 1995).

El criterio de ajuste de curvas utilizado por Aaslid y Tiecks es el menor error estándar de la media de las diferencias entre la curva estimada y la curva real, cuya expresión se muestra en la ecuación 3.5, en donde *V* es la señal de flujo y *n* es el largo de la misma.

|  |  |  |
| --- | --- | --- |
|  |  | *(3.5)* |



*Figura 1‑1. Respuestas al escalón del modelo de A-T usando las 10 tripletas de valores posibles para K, D y T de la tabla 3.1.*

### Modelo Aaslid-Tiecks Decimal.

Este modelo corresponde a una versión mejorada del modelo de *Aaslid-Tiecks* original. En lugar de utilizar la tabla de diez filas, utiliza una tabla con más filas obtenida interpolando los valores de la primera.

Se obtienen los ARI usando señales reales de presión y 91 tripletas de los valores *T*, *D* y *K* (APÉNDICE A), obtenidas mediante interpolación de las 10 originales de *Aaslid-Tiecks* (Nuñez, 2003), con lo que se generan 91 curvas teóricas que representan los distintos niveles del modelo de *Aaslid-Tiecks* Decimal, los que entre el rango de 0 hasta 9, pasando por cada uno de los decimales intermedios. Estas pueden verse representadas en la Figura 3-2.



Figura 1‑2 Modelo de las 91 respuestas teóricas al escalón inverso (Varas, 2013)

## MÁQUINAS DE VECTORES SOPORTE.

### Orígenes.

Durante la década de 1960, en Rusia se desarrolló el algoritmo de vectores soporte basados en una generalización del algoritmo *Generalized Portrait* (Vapnik & Lerner, 1964; Vapnik & Chervonenkis, 1971). Posterior a ello y basados en estos mismos principios, el año 1971 Vapnik y Chervonenkis crearon la llamada teoría VC, en honor a sus creadores, la cual está basada en un método de aprendizaje estadístico y caracteriza propiedades de las máquinas de aprendizaje con las que se puede realizar una buena generalización frente a datos desconocidos (Smola *et al.*, 1998).

En su forma actual, las SVM fueron desarrolladas a principios de los años 90 en los laboratorios AT&T Bell por el mismo Vapnik y sus colaboradores (Boser *et al.*, 1992). En sus inicios, este método sólo era aplicable a problemas simples de clasificación (SVC, *Support Vector Clasification)*, en los cuales la meta es separar los ejemplos pertenecientes a la clase positiva de los pertenecientes a la clase negativa a través de un hiperplano, de forma que el margen de separación entre una y otra sea máximo (Boser *et al.*, 1992).

Esta formulación no disponía de la flexibilidad necesaria como para ser utilizada en problemas más complejos, pero con la introducción del concepto de “margen blando” (*soft margin*) por parte de Cortes y Vapnik (Cortes & Vapnik, 1995), fue posible extender las SVM para trabajar exitosamente con datos ruidosos no separables.

Las posibilidades que presentan las SVM como método de aprendizaje incluyen la clasificación y la regresión. Este último método puede ser usado con recurrencias externas para realizar estimaciones de series de tiempo, y dados los buenos resultados obtenidos al ser aplicados a los problemas de hemodinámica cerebral, (Díaz, 2005; Araya, 2006; Bello, 2007) es este tipo de metodología la que se usa en este trabajo.

### Regresión lineal en las SVM.

El concepto de regresión plasma la necesidad de aproximar o estimar una función a partir de datos disponibles con una determinada precisión, considerando el ruido que puede estar contenido en esta información. Para efectos del estudio que se realiza, se define la *regresión en las máquinas de vectores soporte* (*Support Vector Regression,* SVR).

A partir de una muestra de datos , donde es un vector de *p* dimensiones e es un número real, con *n* el total de datos, se desea encontrar una función lineal de la forma:

|  |  |  |
| --- | --- | --- |
|  |  | * (3.1) |

En esta función, es un vector de dimensión *p*, corresponde al producto punto entre ambos vectores y *b* es un número real. La idea principal es buscar la minimización de la *d*-norma de . Para *d* = 2, este problema se transforma en un problema de optimización cuadrático convexo (Smola, 1996), como se puede apreciar en la ecuación (3.2) sujeto a las restricciones presentadas en (3.3).

|  |  |  |
| --- | --- | --- |
| Minimizar: |  | * (3.2) |
| Sujeto a: |  | * (3.3) |

Se parte de la premisa de que el problema de optimización tiene solución, considerando que existe una función capaz de aproximar todos los datos con precisión *.* Sin embargo, a causa del ruido existente en los datos, dicha función no siempre existe. Debido a ello se deben introducir las denominadas variables de holgura o variables *slack*, las que consideran un pequeño margen de error, ante ciertas condiciones para puntos específicos que se encuentren fuera del espacio de margen máximo. Estas variables se denominan y cuya relación con la precisión es establecida a través de las denominadas *funciones de pérdida* (Vapnik, 1995). Al utilizar la función de pérdida *ε - insensitiva,* donde *ε* > 0, las variables de error se estiman mediante la medición lineal de la distancia *ε* en que se encuentra una muestra, el problema se transforma en el siguiente:

|  |  |  |
| --- | --- | --- |
| Minimizar: |  | (3.4) |
| Sujeto a: |  | (3.5) |

En ellas, la constante *C* > 0 determina el balance entre la influencia del error y la inclinación de la función *f.* La resolución del problema descrito, se presenta en la Figura 3.2 (Smola & Scholkopf, 2004). En la medida en que los puntos se encuentren fuera de la región sombría, contribuyen a los costos, así como las desviaciones son penalizadas de manera lineal. Las condiciones presentadas indican que el problema puede ser resuelto de una manera menos compleja mediante su formulación dual.



*Figura 1‑3. Ajuste de margen suave de pérdida para una SVR lineal.*

##### Formulación dual de la SVR.

La idea clave consiste en construir una función de Lagrange a partir de la función objetivo y sus correspondientes restricciones, mediante la introducción de un conjunto de nuevas variables como se muestra en la ecuación (3.6).

Donde son las denominadas *variables primales*, y   
son los llamados multiplicadores de Lagrange o *variables duales*, los cuales deben ser mayores o iguales a 0. Se presenta un único punto silla que representa la solución óptima, en contraste con las redes neuronales que pueden caer en mínimos locales.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | | (3.6) |
|  |  |  |  |

De acuerdo al *teorema de Kuhn-Tucker,* las derivadas parciales de la ecuación (3.6) con respecto a las variables primales son iguales a cero en el punto silla y en las ecuaciones (3.7 - 3.10) se pueden apreciar las relaciones de las variables duales.

|  |  |  |
| --- | --- | --- |
|  |  | (3.7) |
|  |  | (3.8) |
|  |  | (3.9) |
|  |  | (3.10) |

Finalmente, mediante la sustitución de las ecuaciones presentadas, en la ecuación (3.6), se obtiene el problema dual de la SVR (Smola & Scholkopf, 2004), como se muestra en la nueva función objetivo en la ecuación (3.11), sujeta a las restricciones presentadas en la ecuación (3.12).

|  |  |  |
| --- | --- | --- |
| Maximizar: |  | (3.11) |
| Sujeto a: |  | (3.12) |

En la obtención de la ecuación (3.11) se han eliminado las variables duales y de acuerdo a las ecuaciones (3.9) y (3.11), ya que estas variables no aparecen más en la función objetivo dual, pero están presentes solamente en las condiciones de factibilidad dual. La ecuación (3.7) se puede reescribir como sigue:

|  |  |  |
| --- | --- | --- |
|  |  | (3.13) |

y por lo tanto,

|  |  |  |
| --- | --- | --- |
|  |  | (3.14) |

Ésta es la así llamada expansión de vectores soporte, es decir, puede ser descrito totalmente como una combinación lineal de las muestras de entrenamiento . En cierto modo, la complejidad de la representación de una función es independiente de la dimensionalidad del espacio de entrada , y depende solamente del número de vectores soporte (Smola & Scholkopf, 2004).

##### Vectores Soporte.

Basados en las condiciones de *Karush-Kuhn-Tucker*, se determina que en la solución óptima, el producto de las variables duales y las restricciones presentadas en la ecuación (3.5), deben ser iguales a cero como se presenta en la ecuación (3.14) (Smola & Scholkopf, 2004)

|  |  |  |
| --- | --- | --- |
|  |  | (3.14) |
|  |  | (3.15) |

Lo anterior implica que sólo los ejemplos con su correspondiente yacen fuera del tubo *ε -insensitivo* alrededor de *f.* En segundo lugar no pueden existir de manera simultánea un conjunto de variables duales distintas de cero.

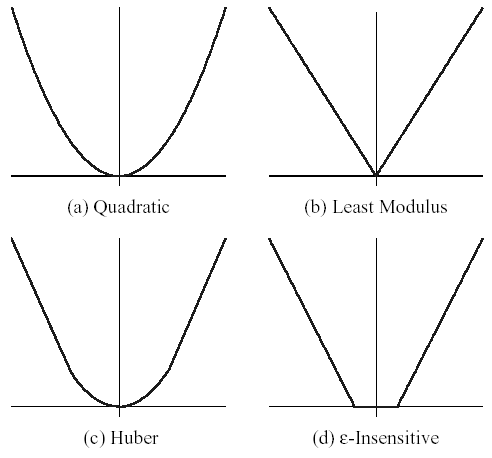
Finalmente para aquellas muestras en que , la variable es igual a cero, al igual que el segundo factor en (3.15). Con estas condiciones es posible el cálculo de *b*, tal como se muestra en la ecuación (3.16).

|  |  |  |
| --- | --- | --- |
|  | para  para | (3.16) |

Al tener todos los elementos involucrados en la regresión se observa que es necesario obtener sólo aquellas muestras cuyos son mayores que cero ya que con ellas basta para la resolución del problema, por lo tanto ya no son necesarias todas los para describir el problema como en el caso de las redes neuronales. Aquellas muestras necesarias son los llamados v*ectores soporte* (VS).

### Funciones de pérdida.

Complementando los puntos anteriores, existen diferentes funciones de pérdida que permiten dar solución a los problemas de regresión, la restricción que se debe cumplir radica en que el problema resultante sea convexo, para asegurar la existencia de una solución única, en este caso, el óptimo. En la Figura 3.3 pueden apreciarse las funciones de pérdida más utilizadas (Gunn, 1998).



(d)

(c)

(b)

(a)

*Figura 1‑4. Gráficos de las funciones de pérdida (a) Cuadrática, (b) Laplaciana,   
(c) Huber y (d) ε -Insensitiva*

### Regresión no lineal.

Dentro de las diferentes fenomenologías estudiadas mediante el uso de SVR, es raro encontrar casos en que solamente las regresiones lineales sean suficientes para capturar el fenómeno. Para dar solución a este problema, se realiza una transformación de los datos de entrada, mediante el uso de una función, a un espacio de características de mayor dimensión, en la cual pueda encontrarse la perspectiva que permita la regresión lineal del problema.

Para la realización de esta transformación, basados en la dependencia de los productos punto entre los datos de entrada, se utilizan las llamadas funciones *Kernel,* representadas por de manera general, lo que implica que el conocimiento explicito de Φ no sea necesario del todo.

### Funciones Kernel.

La condición fundamental que garantiza la utilidad de la función para ser usada como *Kernel* para las transformaciones, es la condición de Mercer*.* Esta condición viene dada por el teorema creado en 1909 en el cual se estipula que una función es *Kernel* (Burges, 1998)*,* si y sólo si, para cualquier la ecuación (3.17) da como resultado un número finito y (3.18) se cumple.

|  |  |  |
| --- | --- | --- |
|  |  | * (3.17) |
|  |  | * (3.18) |

No se puede dejar de mencionar que otra de las condiciones de las funciones *Kernel* es su simetría, es decir, que este tipo de funciones correspondan a un producto punto en algún espacio de mayor dimensión al espacio de entrada y que sea igual a (Betancourt, 2005).

Existen diversas funciones *Kernel* que permiten tratar los distintos problemas considerando variados enfoques. La más utilizada y que ha tenido mejores resultados en estudios similares al realizado en este trabajo, es la llamada función *Kernel de base radial (Radial Basis Function, RBF).* En la Tabla 3.2, se presenta un breve listado de las funciones más utilizadas para realizar la regresión, la función *Kernel lineal*, correspondiente a la formulación lineal de las SVM, por otro lado está la función *Kernel polinomial*, la cual realiza un mapeo polinomial para modelamientos no lineales, se observa además la funcion *Kernel RBF*, mencionado anteriormente y finalmente se aprecia la *Sigmoidal*, similar a la utilizada en las redes neuronales artificiales, sin embargo ésta, presenta la desventaja de no satisfacer la condición de Mercer para todo el margen de parámetros que ella utiliza (Burges, 1999).

*Tabla 3.2. Funciones Kernel más utilizadas*

|  |  |
| --- | --- |
| Nombre | Ecuación |
| Lineal |  |
| Polinomial (grado m) |  |
| RBF |  |
| Sigmoidal |  |

### ν -SVR.

Uno de los grandes problemas que se presenta en las SVR es el alto nivel de parametrización que ésta necesita para obtener buenos resultados. Sin embargo, de no ser por ello los resultados no siempre serían los esperados y en muchos casos el modelamiento de problemas y la acertada representación de los fenómenos serían imposibles.

Al utilizar la función de pérdida *ε-*insensitiva, aparece el problema de escoger un parámetro *ε* adecuado para obtener resultados aceptables.

Según Smola *et al.* (1998) existe una dependencia lineal entre el nivel de ruido y el valor óptimo de *ε*. Sin embargo, esto requiere conocer el modelo de ruido que afecta a los datos, lo cual generalmente es desconocido. Para no hacer frente a este problema, las *ν* - SVR, aparecen como una manera de controlar aún más el comportamiento de las SVR, lo que permite reemplazar *ε* por un parámetro *ν*, el cual lo minimiza automáticamente y permite mantener control sobre la cantidad de vectores soporte y la fracción de errores de entrenamiento, ya que *ν* es igual a la proporción de ambos (Scholkopf *et al.*, 2000). Se tiene, entonces el parámetro es el límite superior para la fracción de errores de entrenamiento y el límite inferior para la fracción de vectores soporte (Hsu *et al.*, 2003).