

# Application of Asymmetric Logarithmic Regression Method, Logit Kernel Regression, and Support Vector Machine for Predicting the Satisfaction Level of Clients of Telecommunication Companies

Adama Esperilla Jhonny Ivan, Loyola Perez Jessica Ruth, Melgarejo Huaman Walter Joel

Asignatura: Métodos Económicos I

Escuela Profesional de Ingeniería Estadística

Facultad de Ingeniería Económica Estadística y Ciencias Sociales

Universidad Nacional de Ingeniería

## Resumen

El presente trabajo desarrolla una comparación entre 3 metodologías para determinar el modelo con mayor poder predictivo en la satisfacción de los clientes para la empresa TELCO con respecto a la cobertura de red. Las metodologías abordadas son: “**Modelos de regresión logística asimétrica**” (Cloglog). La segunda metodología es “**Regresión logística de Kernel**”, en esta se empleó la transformación de Kernel y finalmente la “**Maquina vectores de Soporte**”, se comparó estos modelos predictivos teniendo como resultado final que el modelo de regresión logística kernel tiene mejor capacidad predictiva.

## INTRODUCCIÓN

El modelo de regresión de kernel logístico esta denominado como “el modelo del futuro”, las investigaciones de este modelo son recientes ya que es nuevo en el mercado y aun más, es nuevo para nuestro país.

Por ello nuestra investigación pone en conocimiento la importancia de este modelo estadístico y su aplicación en el mercado, aplicado a la satisfacción de los clientes con respecto a la cobertura de red de la empresa TELCO. Adicional a la investigación se realizara una comparación del modelo kernel logístico con regresión logística (cloglog) y SVM.

Ha habido numerosas aplicaciones relativamente recientes y las investigaciones sobre el modelo, incluyendo Bekhor et al. (2002), Bhat (1997 y 1998), Bolduc, Fortin y Fournier (1996), Brownstone, Bunch y tren (2000), Brownstone y Train (1999), Goett, Hudson, y el tren (2000), Gönül y Srinivasan (1993), Greene (2000), Mehndiratta y Hansen (1997).

El objetivo de este documento es la comparación y la capacidad predictiva que tendrá el kernel logístico con respecto a la regresión logística cloglog y la Maquina de Vectores Soporte.

## PROBLEMA GENERAL

¿Cuál es la capacidad predictiva de los modelos logit kernel en comparación con el modelo de regresión logística asimétrica y SVM para la satisfacción de los clientes de la empresa TELCO con respecto a la cobertura de red para el año 2013?

### Problemas específicos

¿Qué modelo proporciona mejores resultados para la predicción de la satisfacción de clientes con respecto a la cobertura de red para la empresa TELCO para el año 2013 en la muestra tomada?

¿Qué modelo proporciona mejores resultados para la predicción de la satisfacción de clientes con respecto a la cobertura de red para la empresa TELCO para el año 2013 haciendo una inferencia para la población?

## OBJETIVOS

### Objetivo general

Comparar la eficiencia del modelo de Regresión Logística Asimétrica, Modelo de Regresión Logit Kernel y Máquina de Vectores de Soporte para la predicción de la satisfacción de clientes de la empresa TELCO con respecto a la cobertura de red en el año 2013.

### Objetivos específicos

Obtener un modelo predictivo, que nos proporcione los resultados más adecuados con respecto a la cobertura de red para los clientes de la empresa TELCO en el año 2013 en la muestra.

Obtener un modelo predictivo, que nos proporcione los resultados más adecuados con respecto a la cobertura de red para los clientes de la empresa TELCO en el año 2013 haciendo una inferencia para la población.

## DESCRIPCIÓN DE LA SOLUCIÓN

### KERNEL LOGÍSTICO.

Es un modelo de elección discreta que tiene ambos trastornos como probit y un aditivo independiente y valor

distribuido idénticamente extrema (o Gumbel) perturbación a la logit multinomial. El resultado es un modelo intuitivo, práctico y de gran alcance que combina la flexibilidad de probit con la docilidad de logit. Por esa razón logit mixto (logit kernel) ha sido considerado el "modelo del futuro".

La metodología es la siguiente:

$$\text{logit}(y(x)) = w \cdot \varphi(x) + b \quad \dots (1)$$

w: es un vector de parámetros del modelo.

$\varphi(x)$ : representa una transformación no lineal a los vectores de entrada.

$$y(x) = \frac{1}{1 + e^{-w \cdot \varphi(x) - b}} \quad \dots (2)$$

La función logit que contiene los valores de la probabilidad de satisfacción del cliente en el rango [0,1].

Sea:  $K(x, x') = \varphi(x) \cdot \varphi(x')$

La transformación no lineal es llamada como la función Kernel que es el producto escalar entre las imágenes de las transformaciones no lineales.

Se sabe que existen diferentes funciones Kernel, hemos utilizado: Kernel Gaussiana (Función de Base Radial):

$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$ ; donde  $\sigma$  es un parámetro de inflexión.

Para nuestro modelo utilizamos la función kernel de base radial ya que es la que normalmente más se ajusta y da mejores resultados.

El vector óptimo del parámetro (w) que puede ser hallado con la minimización de la función de costo y puede ser escrito en una combinación lineal como lo siguiente:

$$w = \sum_{i=1}^n \alpha_i \cdot \varphi(x) \quad \dots (3)$$

La forma final de la regresión logística kernel simétrico es:

$$\text{logit}(y(x)) = \sum_{i=1}^n \alpha_i K(x_i, x) + b \quad \dots (4)$$

Donde,  $\alpha$  es un vector de parámetros del modelo dual, los mejores parámetros de inflexión son 0.02 y 0.025 respectivamente.

### Máquina de Vectores Soporte

(BETANCOURT, 2005) Una Máquina de Soporte Vectorial (SVM) aprende la superficie decisión de dos clases distintas de los puntos de entrada. Como un clasificador de una sola clase, la descripción dada por los datos de los vectores de soporte es capaz de formar una frontera de decisión alrededor del dominio de los datos de aprendizaje con muy poco o ningún conocimiento de los datos fuera de esta frontera. Los datos son

mapeados por medio de un KERNEL Gaussiano u otro tipo de KERNEL a un espacio de características en un espacio dimensional más alto, donde se busca la máxima separación entre clases. Esta función de frontera, cuando es traída de regreso al espacio de entrada, puede separar los datos en todas las clases distintas, cada una formando un agrupamiento.

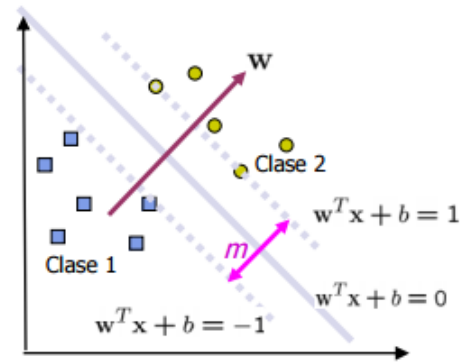


Figura 01. La frontera de decisión debe estar tan lejos de los datos de ambas clases.

Del SVM hay dos tipos: los linealmente separables y los linealmente no separables, para el análisis del trabajo se realizara el análisis lineal no separable, aplicando la transformación gaussiana para hacer una mejor clasificación (demostración ver anexo).

Truco del Kernel para el caso no linealmente separable. Hay una buena propiedad de la SVM la cual es que no es necesario tener ningún conocimiento. Nosotros sólo necesitamos una función  $K(x_i, x)$  llamada kernel que calcule el producto punto de los puntos de entrada en el espacio de características Z, esto es

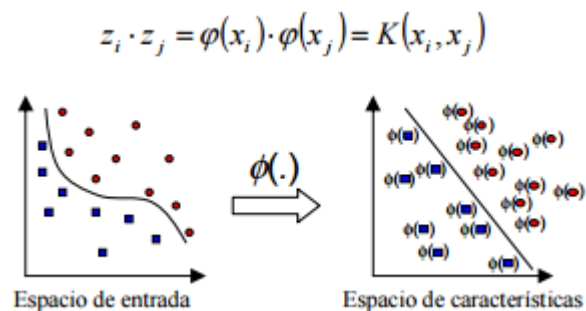


Figura 02. Idea del uso de un kernel para la transformación del espacio de los datos

### Regresión logística asimétrica

La regresión logística (RL) forma parte del conjunto de métodos estadísticos que caen bajo tal denominación y es la variante que corresponde al caso en que se valora la contribución de diferentes factores en la ocurrencia de un evento simple. En general, la regresión logística es adecuada cuando la variable de respuesta Y es politómica (admite varias categorías de respuesta, tales como mejora mucho,

empeora, se mantiene, mejora, mejora mucho), pero es especialmente útil en particular cuando solo hay dos posibles respuestas (cuando la variable de respuesta es dicotómica), que es el caso más común. La RL es una de las técnicas estadístico inferenciales más empleadas en la producción científica contemporánea. Surge en la década del 60, su generalización dependía de la solución que se diera al problema de la estimación de los coeficientes. El algoritmo de Walker-Duncan para la obtención de los estimadores de máxima verosimilitud vino a solucionar en parte este problema, pero era de naturaleza tal que el uso de computadoras era imprescindible.

Debido a los problemas del modelo de probabilidad lineal, se han buscado modelos alternativos de la forma.

$$Y = F(\alpha + \beta x) + \epsilon(x) \quad \dots (5)$$

Con  $\epsilon(x)$  vv.aa independientes con esperanza 0, con lo que el modelo sobre la probabilidad condicionada se puede escribir como.

$$p(x) = F(\alpha + \beta x) \quad \dots (6)$$

Con F función monótona creciente. También se puede expresar como sigue

$$F^{-1}(p(x)) = (\alpha + \beta x) \quad \dots (7)$$

Es decir, se busca una función F cuya inversa transforme las probabilidades condicionadas  $p(x)$  y posteriormente, modelar linealmente esta transformación. Según se elija una determinada función F se tienen distintas formulaciones.

**Transformación cloglog.** Esta transformación es de la forma

$$p(x) = 1 - \exp[-\exp(\alpha + \beta x)] \quad \dots (8)$$

En la forma lineal sería

$$\log[-\log(1 - p(x))] = \alpha + \beta x \quad \dots (9)$$

Esta transformación no tiene un comportamiento simétrico, sino que se aleja del valor de probabilidad 1 de forma más rápida de lo que se acerca al valor 0.

### Intervalos e hipótesis para la diferencia de proporciones de dos muestras independientes

Los límites para el intervalo de una diferencia de proporciones correspondientes a dos muestras independientes son:

$$\widehat{p}_1 - \widehat{p}_2 \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\widehat{p}_1(1 - \widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1 - \widehat{p}_2)}{n_2}} \quad \dots (10)$$

Donde el símbolo  $Z_{\frac{\alpha}{2}}$  es el mismo valor crítico que antes,  $\text{prob}(Z > Z_{\frac{\alpha}{2}}) = \alpha/2$ , y corresponde a un intervalo de confianza  $1 - \alpha\%$

Prueba de hipótesis:

$$Z_c = \frac{(\widehat{p}_1 - \widehat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\widehat{p}_1(1 - \widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1 - \widehat{p}_2)}{n_2}}} \quad \dots (11)$$

Para:

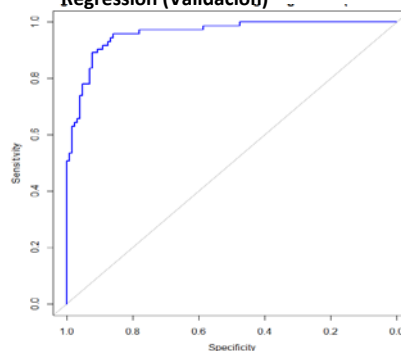
$$H_0: (\widehat{p}_1 - \widehat{p}_2) > 0.05$$

Se rechaza la hipótesis nula si  $-Z_c < -Z_{\alpha/2}$

## RESULTADOS

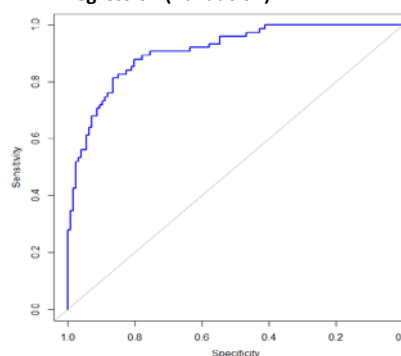
### COMPARACION POR CURVAS ROC DE LOS MODELOS (KLR, SVM Y RCLOGLOG)

Figura 03. Curva ROC del Kernel Logit Regression (Validación)



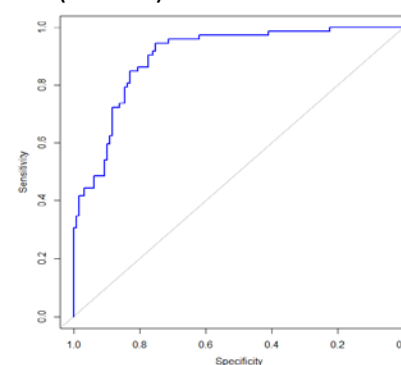
AUC=0.9602

Figura 04. Curva ROC de Cloglog Regression (Validación)



AUC=0.9083

Figura 05. Curva ROC de SVM (validación)



AUC=0.8997

De las figuras (3, 4 y 5) se tiene que el Área bajo la curva es mayor en el modelo KRL con un 96.02%, seguido por el CLOGLOG con un 90.83% y por último, SVM con un 89,97%

**Cuadro 01. Validación del Modelo Usando Medidas de Evaluación Estadísticas**

N o.	Parámetros	Muestra de Validación		
		KLR	CLOGLOG	SVM
1	Verdaderos Positivos	57	46	43
2	Verdaderos Negativos	120	118	115
3	Falsos Positivos	16	29	29
4	Falsos Negativos	8	8	14
5	Valor Predictivo Positivo	78.08%	61.33%	59.72%
6	Valor Predictivo Negativo	93.75%	93.65%	89.15%
7	Sensibilidad	87.69%	85.19%	75.44%
8	Especificidad	88.24%	80.27%	79.86%
9	Exactitud	88.06%	81.59%	78.61%

Del cuadro 1: muestra la validación del modelo usando medidas de evaluación estadística, para los modelos KLR (RBF), CLOGLOG Y SVM.

En la comparación de la validación, la diferencia entre la exactitud de la muestra de entrenamiento y validación es menor que el 5% entonces el modelo es eficiente, así mismo, tanto en la sensibilidad y especificidad la diferencia es menor también. De la comparación el que tiene mayor exactitud y sensibilidad es el KRL (RBF) ya que su exactitud es del 86.89% y sensibilidad del 83.39%, por lo tanto con estos indicadores el modelo KRL tiene mejor predicción que los modelos CLOGLOG Y SVM Del análisis anterior se utilizaron los porcentajes de aciertos que tuvo el modelo en la predicción, pero para tener un mejor análisis utilizaremos la curva COR (ROC) para la comparación de los modelos de validación (ver figuras 3, 4 y 5).

### ANALIZANDO MEDIANTE PRUEBA DE HIPÓTESIS E INTERVALOS DE CONFIANZA

Ho:  $p_1 - p_2 \geq 0.05$

**Cuadro 02. Prueba de hipótesis de la diferencia de proporciones en muestras desbalanceadas**

Modelos e Indicadores	p1	p2	n1	n2	Zc	Zt	Conclusión
KLR	83%	88%	801	201	-3.49	-1.64	Se rechaza la Ho
CLOGLOG	82%	85%	801	201	-2.73	-1.64	
SVM	78%	75%	801	201	-0.845	-1.64	No se rechaza la Ho

Se define por experiencia de las encuestadoras que la diferencia entre las sensibilidades de las muestras de validación y muestras de entrenamiento debe ser menor al 5 %, pero como esta es una muestra lo validamos con una prueba de hipótesis para asegurar con una confiabilidad del 95% que existe evidencia estadística de que hay una diferencia menor al 5% en los modelos KLR y CLOGLOG.

**Cuadro 03. Intervalo de confianza para la diferencia de proporciones**

Intervalos de Confianza para los Indicadores de Desempeño del Modelo		MUESTRA DE VALIDACIÓN		
		KLR-CLOGLOG	CLOGLOG-SVM	KLR-SVM
Sensibilidad	límite superior	0.055	0.132	0.156
	límite inferior	-0.005	0.063	0.089

Se aprecia en el cuadro 3 que con un nivel de confianza del 95% no podemos afirmar que el KLR es mayor que el CLOGLOG ya que en el intervalo de confianza poblacional incluye el valor 0, por ello son estadísticamente iguales a nivel de población, por lo tanto no podemos afirmar que la sensibilidad es mayor que otra.

### CONCLUSIONES

Al comparar los modelos a nivel de la muestra analizamos el AUC de los modelos y se obtuvo los siguientes resultados: KRL =96.02%, SVM=90.83% y RL(cloglog)= 89.97%, estos porcentajes representan que tan eficientes son los modelos (en la muestra de validación).

Al analizar los modelos, el KRL obtuvo el mayor porcentaje de AUC mostrado en los graficos(3, 4 y 5) además de tener mejor precisión tanto en la sensibilidad y exactitud (cuadro 1) en comparación con los modelos de regresión (cloglog) y SVM solo para este caso, satisfacción de clientes de la empresa TELCO en la muestra de validación.

Usando la prueba de hipótesis para saber si la diferencia en la población de la sensibilidad en la muestra de entrenamiento y validación será menor a 5% con una confiabilidad del 95%, observamos que en el cuadro 2 tanto el KLR como la Regresión Logística CLOGLOG rechazan la  $H_0: p_1-p_2 \geq 0.05$ , se puede decir que la predicción de estos dos modelos son confiables a un 95%, y observamos en el cuadro 1 que la sensibilidad del KLR es mayor que el CLOGLOG pero en los resultados del cuadro 3, en el intervalo de confianza de la diferencia de sensibilidades con una confiabilidad del 95% se dice que estadísticamente son iguales ya que la diferencia incluye el valor 0.

### BIBLIOGRAFÍA

- [1] Nancy Dávila, Dolores García Artilles, José Pérez Sánchez, Emilio Gómez Déniz. *An Asymmetric Logit Model to explain the likelihood of success in academic results* Universidad de Las Palmas de Gran Canaria, Universidad de Granada, 2015.
- [2] Esther Gutiérrez Alonso, *Aplicación De Las Máquinas De Soporte Vectorial Para El Reconocimiento De Matrículas*, Madrid, 2015.
- [3] Joan Walker, Moshe Ben Akiva, Denis Bolduc, *Identification of the Logit Kernel (or Mixed Logit) Model*, USA, 2003
- [4] José Luis Cañadas Reche, *Regresión logística. Tratamiento computacional con R.*, Universidad de Granada España, 2013

### ANEXOS

#### Caso no linealmente separable

Si el conjunto S no es linealmente separable, violaciones a la clasificación deben ser permitidas en la formulación de la SVM.

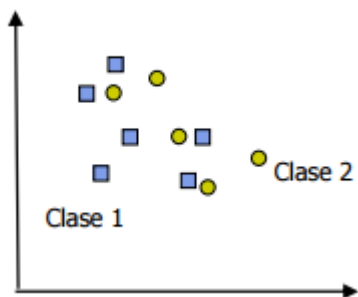


Figura 06. Caso no linealmente separable

Para tratar con datos que no son linealmente separables, el análisis previo puede ser generalizado introduciendo algunas variables no-negativas  $\xi_i \geq 0$  de tal modo que es modificado a

$$y_i(w \cdot z_i + b) \geq 1 - \xi_i, \quad i = 1, \dots, l \quad (12)$$

Los  $\xi_i \neq 0$  en (1) son aquellos para los cuales el punto  $x_i$  no satisface la respuesta. Entonces el término  $\sum_{i=1}^l \xi_i$  puede ser tomado como algún tipo de medida del error en la clasificación. El problema del hiperplano óptimo es entonces redefinido como la solución al problema

$$\min \left\{ \frac{1}{2} w \cdot w + C \sum_{i=1}^l \xi_i \right\}$$

$$s. a \quad y_i(w \cdot z_i + b) \geq 1 - \xi_i \quad i = 1, \dots, l \quad (13)$$

$$\xi_i \geq 0, \quad i = 1, \dots, l$$

Donde C es una constante. El parámetro C puede ser definido como un parámetro de regularización. Este es el único parámetro libre de ser ajustado en la formulación de la SVM. El ajuste de éste parámetro puede hacer un balance entre la maximización del margen y la violación a la clasificación. Más detalles se pueden encontrar en [12]. Buscando el hiperplano óptimo en (13) es un problema QP, que puede ser resuelto construyendo un Lagrangiano y transformándolo en el dual

$$\text{Max } w(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j z_i \cdot z_j$$

$$s. a \quad \sum_{i=1}^l y_i \alpha_i = 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, \dots, l \quad (14)$$

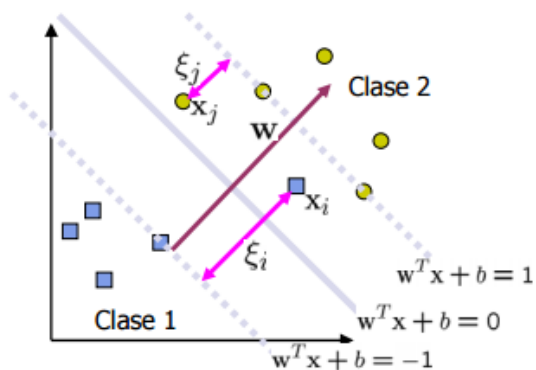
Donde  $\alpha = (\alpha_1 \dots \alpha_l)$ , es un vector de multiplicadores de Lagrange positivos asociados con las constantes en (12). El teorema de Kuhn-Tucker juega un papel importante en la teoría de las SVM. De acuerdo a este teorema, la solución  $\alpha_i$  del problema (14) satisface:

$$\overline{\alpha}_i (Y_i(\overline{w} \cdot z_i + \overline{b}) - 1 + \overline{\xi}_i) = 0, \quad i = 1, \dots, l \quad (15)$$

$$(C - \overline{\alpha}_i) \overline{\xi}_i = 0 \quad i = 1, \dots, l \quad (16)$$

De esta igualdad se deduce que los únicos valores  $\alpha_i \neq 0$  (16) son aquellos que para las constantes en (1) son satisfechas con el signo de igualdad. El punto  $x_i$  correspondiente con  $\alpha_i > 0$  es llamado vector de soporte. Pero hay dos tipos de vectores de soporte en un caso no separable. En el caso  $0 < \alpha_i < C$ , el correspondiente vector de soporte  $x_i$  satisface las igualdades  $y_i(w \cdot z_i + b) = 1$  y  $\xi_i = 0$ . En el caso  $\alpha_i = C$ , el correspondiente  $\xi_i$  es diferente de cero y el correspondiente vector de soporte  $x_i$  no satisface la respuesta. Nos referimos a estos vectores de soporte como errores. El punto  $x_i$  correspondiente con  $\alpha_i = 0$

es clasificado correctamente y está claramente alejado del margen de decisión



**Figura 07.** Aparición del parámetro de error  $\xi_i$  en el error de clasificación

Para construir el hiperplano óptimo  $w \cdot z + b$ , se utiliza

$$\bar{w} = \sum_{i=1}^l \bar{\alpha}_i y_i z_i \quad (17)$$

Y el escalar  $b$  puede ser determinado de las condiciones de Kuhn-Tucker (16). La función de decisión generalizada es tal que

$$f(x) = \text{sign}(w \cdot z + b) = \text{sign}\left(\sum_{i=1}^l \alpha_i y_i \cdot z + b\right) \quad (18)$$

Kernel más usados:

Lineal:  $K(x_i, x_j) = x_i^t \cdot x_j$

Gaussiana:  $K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$

Polinómico:  $K(x_i, x_j) = (x_i^t \cdot x_j + 1)^n$