

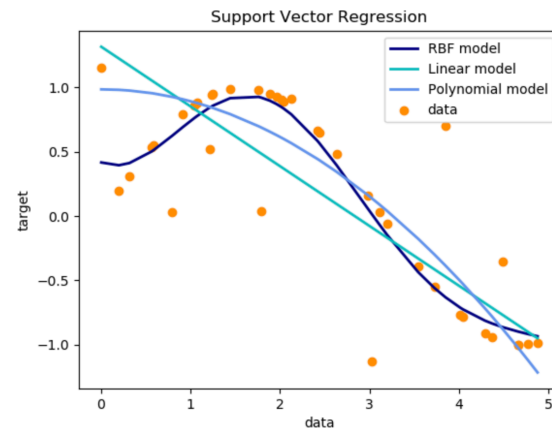
# Regresión con Support Vector Machine

# SVM para Regresión

- Las Máquinas de Soporte Vectorial sirven tanto para regresiones lineales como no lineales, por eso las llamamos SVR
- En lugar de ajustar el mayor corredor (o la mayor calle) posible entre dos clases, ajustando el margen de los mismos como hacen en general las SVM, las SVR intentan mantener cuántas más observaciones posibles del conjunto de datos dentro del corredor en torno a la recta limitando unos márgenes máximos.
- La anchura del pasillo (o la calle) en torno a la recta se controla mediante un hiper parámetro,  $\epsilon$ .

# SVM para Regresión

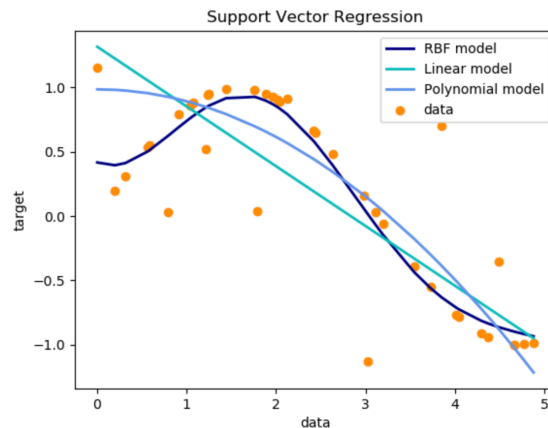
- La SVR lleva a cabo una regresión lineal en un espacio de dimensión superior.
- Podemos pensar que al crear una SVR es como si cada punto de datos del conjunto de entrenamiento representara su propia dimensión. Al evaluar el núcleo entre un punto de test y un punto del conjunto de entrenamiento, el resultado es la coordenada del punto de test en dicha dimensión.



[http://scikit-learn.org/stable/  
auto\\_examples/svm/  
plot\\_svm\\_regression.html#](http://scikit-learn.org/stable/auto_examples/svm/plot_svm_regression.html#)

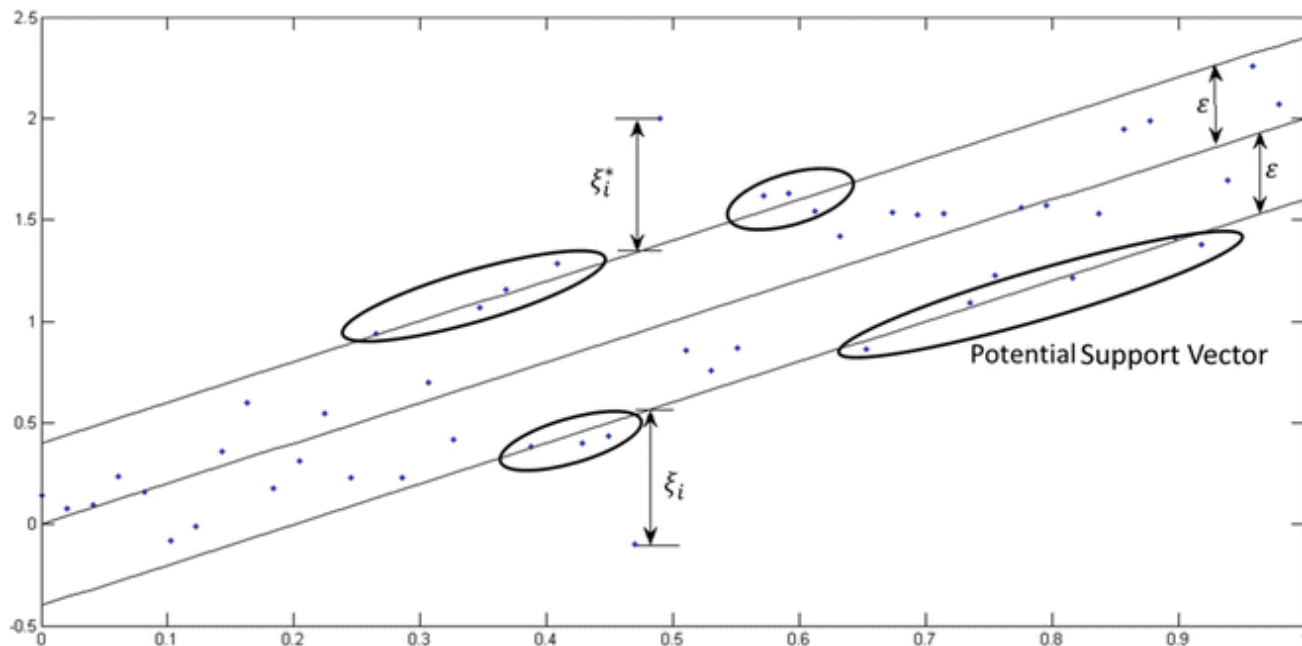
# SVM para Regresión

- El vector que obtenemos al evaluar el punto de test para todos los puntos del dataset,  $k$  es la representación del punto de test en la dimensión superior.
- Cuando tenemos ese vector en el espacio de dimensión superior, lo utilizamos para hacer la regresión propiamente dicha.



[http://scikit-learn.org/stable/  
auto\\_examples/svm/  
plot\\_svm\\_regression.html#](http://scikit-learn.org/stable/auto_examples/svm/plot_svm_regression.html#)

# SVM para Regresión



[https://link.springer.com/chapter/10.1007/978-1-4302-5990-9\\_4](https://link.springer.com/chapter/10.1007/978-1-4302-5990-9_4)

# SVM para Regresión

Se necesita que el conjunto de entrenamiento  $\mathcal{T} = \{\vec{X}, \vec{Y}\}$  cubra el dominio de interés y vaya acompañado de las soluciones en dicho dominio.

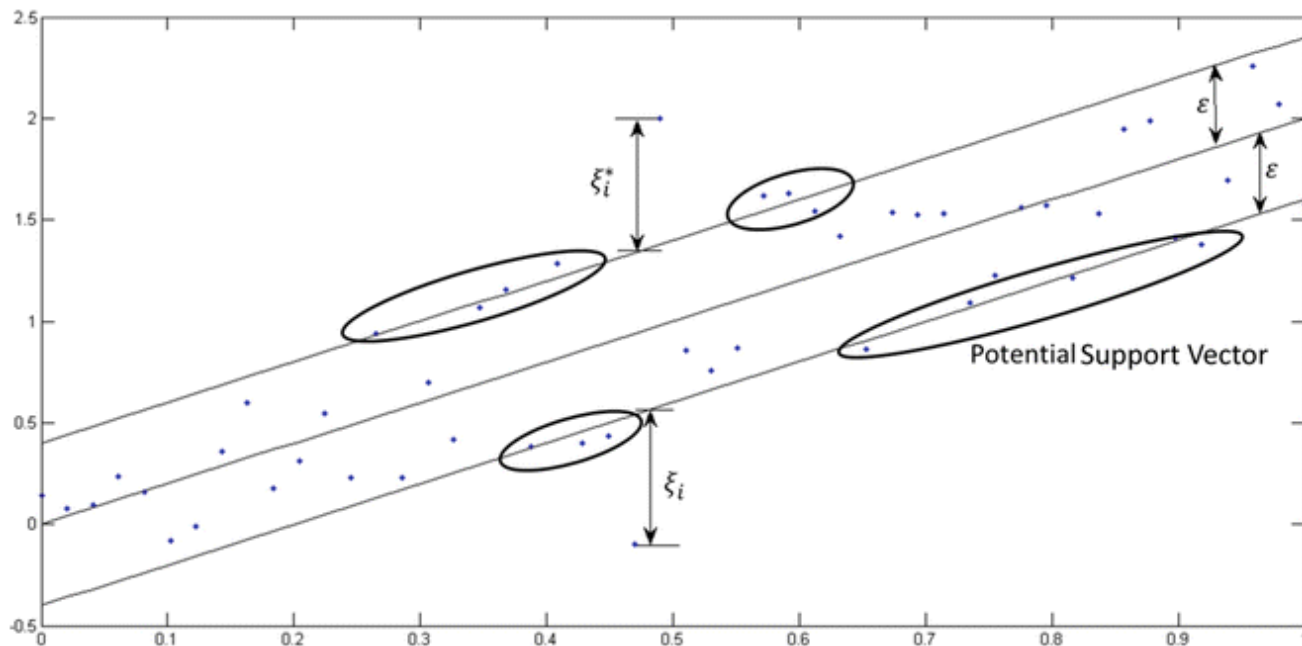
El trabajo de la SVM es aproximar la función que se ha usado para generar los puntos del conjunto de entrenamiento:  $F(\vec{X}) = \vec{Y}$

# SVM para Regresión

En un problema de clasificación, los vectores  $X$  se utilizan para definir un hiperplano que separe las dos categorías en nuestra solución.

Estos vectores se utilizan para llevar a cabo la regresión lineal. Los vectores más cercanos al punto de test se llaman vectores de soporte. Podemos evaluar nuestra función en cualquier lugar, por lo que cualquier vector podría estar más cerca de nuestra ubicación de evaluación de prueba.

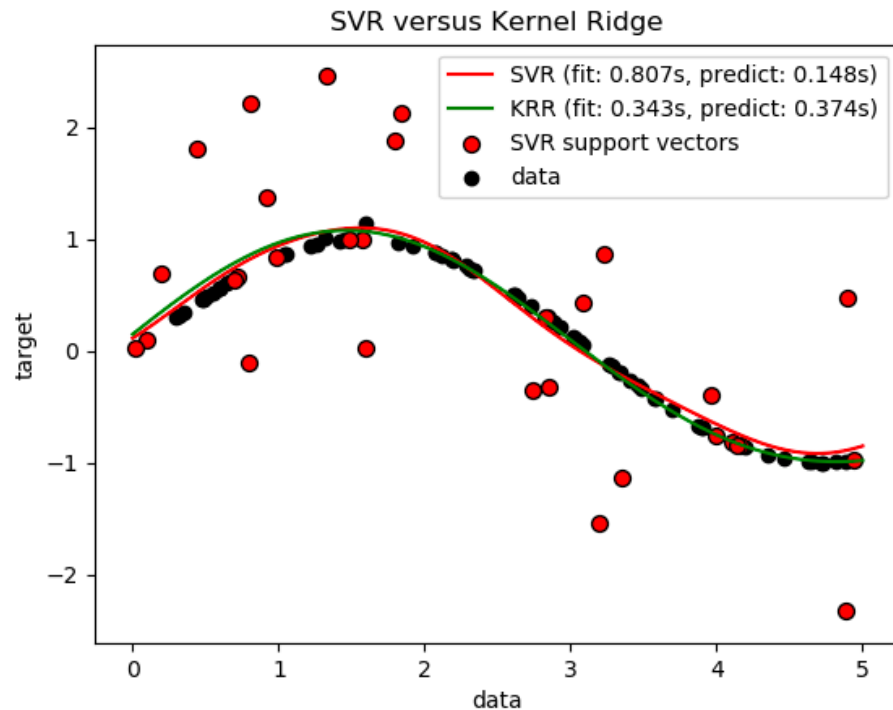
# SVM para Regresión



[https://link.springer.com/chapter/10.1007/978-1-4302-5990-9\\_4](https://link.springer.com/chapter/10.1007/978-1-4302-5990-9_4)



# ¿Cómo funciona un SVR?



# Construir un modelo de SVR

1. Tener un conjunto de entrenamiento,  $\mathcal{T} = \{\vec{X}, \vec{Y}\}$
2. Elegir un núcleo y sus parámetros así como llevar a cabo cualquier regularización que sea necesaria
3. Crear la matriz de correlaciones,  $K$
4. Entrenar el modelo, de forma exacta o aproximada para obtener los coeficientes de contracción,  $\vec{\alpha} = \{\alpha_i\}$
5. Utilizar estos coeficientes para crear un estimador

$$f(\vec{X}, \vec{\alpha}, x^*) = y^*$$

# El núcleo de SVM

---

Lo siguiente es elegir un núcleo:

- Lineal  $\langle x, y \rangle$
- No Lineal  $\langle \varphi(x), \varphi(y) \rangle = K(x, y)$ 
  - Gaussiano

Regularización

- Ruido

# Matriz de Correlaciones

---

Matriz de Correlaciones

$$K_{i,j} = \exp \left( \sum_k \theta_k |x_k^i - x_k^j|^2 \right) + \varepsilon \delta_{i,j}$$

# Paso de optimización

La parte principal del algoritmo es  $K\vec{\alpha} = \vec{y}$

- $\vec{y}$  es el vector de valores del conjunto de entrenamiento
- $K$  es la matrix de correlaciones
- $\vec{\alpha}$  es el conjunto de incógnitas, para las cuales resolvemos el sistema de ecuaciones lineal como

$$\vec{\alpha} = K^{-1}\vec{y}$$

# Vector de correlaciones

- Cuando tenemos la estimación de los parámetros  $\vec{\alpha}$  usamos los coeficientes que hemos calculado durante el paso de optimización y el kernel que habíamos elegido al inicio.
- Para estimar el valor  $y^*$  para un punto de test  $x^*$  calculamos el vector de correlaciones  $\vec{k}$ ,

$$k_i = \exp \left( \sum_k \theta_k |x_k^i - x_k^*|^2 \right)$$

de donde:  $y^* = \vec{\alpha} \cdot \vec{k}$

# En resumen

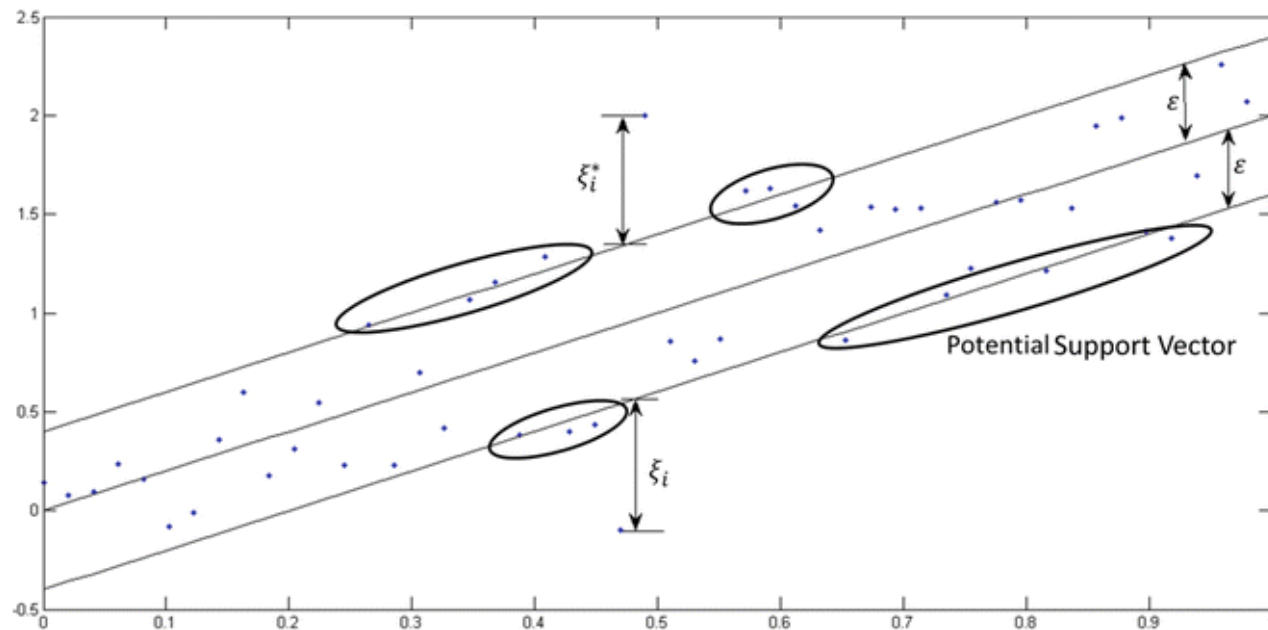
---

Intentemos simplificar las cosas:

La SVR tiene un objetivo de regresión diferente a la regresión lineal.

- En la regresión lineal se intenta minimizar el error entre la predicción y los datos.
- En SVR el objetivo es que los errores no superen el umbral establecido.

# En resumen



[https://link.springer.com/chapter/10.1007/978-1-4302-5990-9\\_4](https://link.springer.com/chapter/10.1007/978-1-4302-5990-9_4)



# Recursos adicionales

---

<https://alex.smola.org/papers/2004/SmoSch04.pdf>

<https://www.mathworks.com/help/stats/understanding-support-vector-machine-regression.html>

<https://stats.stackexchange.com/questions/82044/how-does-support-vector-regression-work-intuitively>

# SVM para Regresión

Los procesos gaussianos son una forma particular de SVM. La diferencia entre los dos radica en la elección del núcleo y la función de pérdida. La forma funcional del kernel determina qué vectores en su conjunto de entrenamiento influyen más fuertemente en la regresión y la forma de su estimador. La elección de la función de pérdida determina los coeficientes utilizados en la regresión. Juntas, estas dos piezas determinan totalmente la forma y la precisión de su estimador. Aunque esto hace que parezca que los dos son totalmente diferentes, el espíritu de los dos es idéntico.