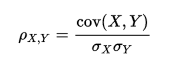
En relación al marco teórico del proyecto, procederemos a documentar detalladamente las tres familias de técnicas de selección de características, detallando su funcionamiento, ventajas e inconvenientes, para posteriormente analizar los casos de uso de cada una, así como una comparación entre estas a fin de mostrar sus diferencias de cara a la elección de uso de cada una en función del contexto de un problema dado.

**Filter methods (métodos de filtrado)**: Se usan principalmente como un método de preprocesamiento, en el cual al grupo de características inicial se le aplica una serie de pruebas estadísticas con el fin de obtener las que mayor puntuación obtengan en estas.



Las pruebas más usadas para filtrar las características son:

* Correlación de Pearson: Cuantifica la dependencia lineal entre dos variables. Su fórmula viene dada por:



Siendo:

X e Y las dos variables.

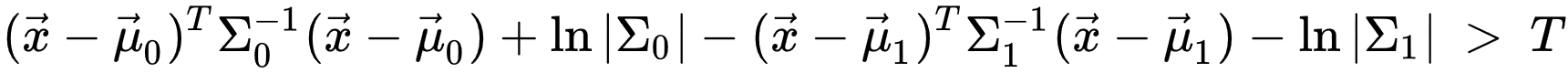
cov(X, Y) la covarianza entre X e Y.

[σ](https://es.wikipedia.org/wiki/%CE%A3)x la desviación estándar de la variable X.

[σ](https://es.wikipedia.org/wiki/%CE%A3)y la desviación estándar de la variable Y.

Este valor estará comprendido en el rango [-1, 1], interpretándose un valor igual a 1 como signo de una relación directa perfecta entre las dos variables, un valor igual a -1 como signo de una relación inversa perfecta, y un valor igual a 0 nos dice que no hay relación entre estas. Los valores comprendidos en el rango (0, 1) indicarán que hay relación directa y los que estén en el rango (-1, 0) indicarán que hay relación inversa.

* Análisis discriminante lineal (ADL): Clasifica observaciones en grupos según sus características, calculando la probabilidad de que pertenezcan a dichos grupos y clasificándola en el que haya obtenido mayor probabilidad. Mediante el uso del teorema de Bayes y siendo el conjunto de observaciones, se asume que las probabilidades de densidad p(|y = 0) y p(|y = 1) siguen una distribución normal con media y covarianza (0, 0) y (1, 1) respectivamente, y siendo T el umbral que define si la observación pertenece a la segunda clase (en caso de superar el umbral):

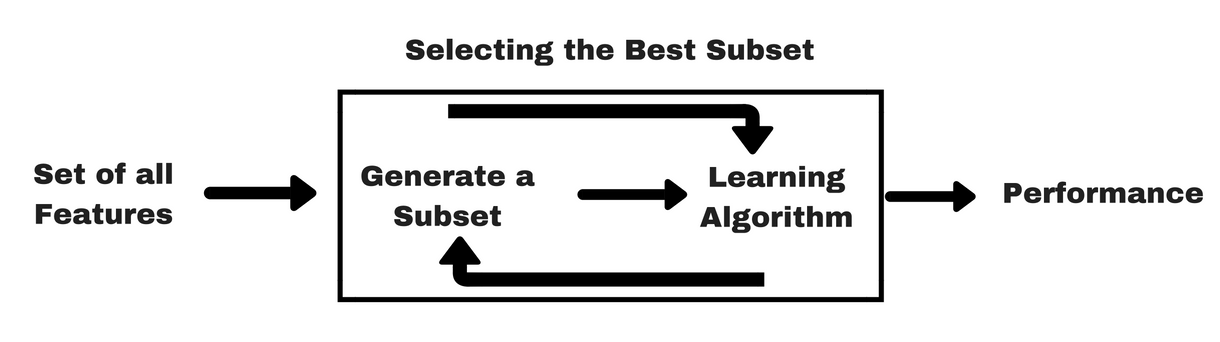


ADL supone que 0=1= y que las covarianzas tienen rango completo, de modo que se cancelan, simplificando los cálculos.

* Análisis de varianza (ANOVA por sus siglas en inglés): Similar al ADL, pero usa una o más características categóricas independientes y una continua, devolviendo un valor estadístico que indica si las medias de varios grupos son iguales.
* Chi-cuadrado: Prueba en la que la medida sigue una distribución x2. Usa la distribución de frecuencias de los grupos de características para evaluar la probabilidad de correlación. Un ejemplo es la prueba de Bartlett, que comprueba si k muestras son de poblaciones con la misma varianza.

Los inconvenientes de usar métodos de filtrado vienen dados al no tenerse en cuenta las relaciones entre las variables, de modo que se obtienen previsiones menos fiables, pero por el contrario son más fáciles de implementar y tienen menor coste computacional.

**Wrapper methods (métodos de envoltura)**: Se usa un subconjunto de características del conjunto inicial para entrenar un modelo, y en base a los resultados de este, se decide si añadir o eliminar características.



Los métodos de envoltura más frecuentes son:

* Selección hacia adelante: Se empieza con un conjunto vacío de características, al que se irá añadiendo la que obtenga una mayor mejora en el modelo hasta que se obtenga una mejora menor que en el paso previo.
* Eliminación hacia atrás: Al contrario que en la selección hacia adelante, se empieza con el conjunto de todas las variables y se va eliminando la menos significativa para el modelo, hasta que no se mejore el rendimiento de la iteración anterior.
* Búsqueda bidireccional: Se elige un método de partida, que puede ser una selección hacia adelante o una eliminación hacia atrás, y se ejecutan las iteraciones al igual que en dichos métodos, con la diferencia de que, si en la próxima iteración se obtiene un rendimiento menos que en la anterior, se vuelve a esta iteración y se realiza seleccionando o eliminando otra característica, respectivamente del método de partida elegido.

Aquí debemos tener en cuenta el criterio de parada establecido para el método elegido, ya que aparte de ser una disminución en el rendimiento en comparación a la iteración anterior, en caso de una búsqueda bidireccional bien podría ser por ejemplo un límite establecido previamente en cuanto al número de características máximas o mínimas del subconjunto que se va obteniendo.

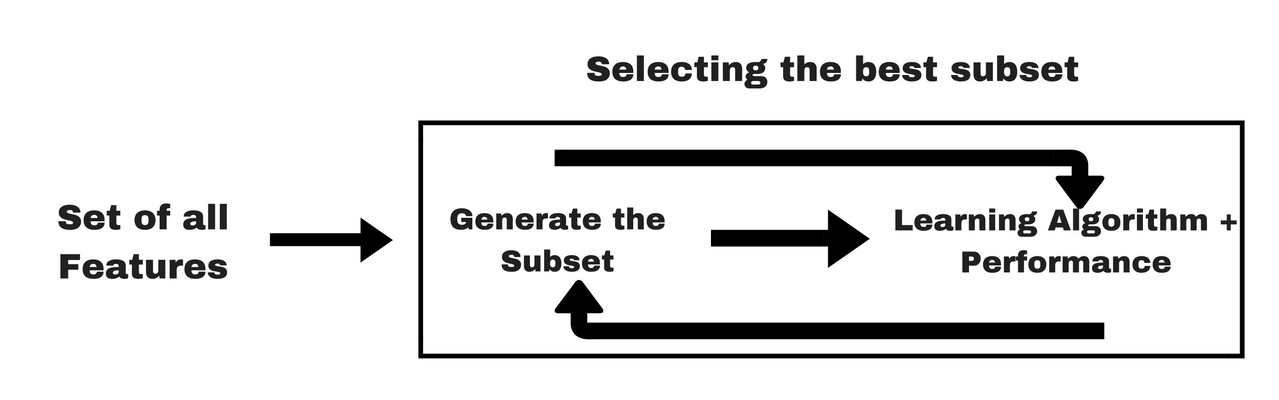
El uso de métodos de envoltura provee de las siguientes ventajas:

* Se tiene en cuenta la interacción entre las variables.
* Obtienen el subconjunto óptimo de características para el algoritmo de aprendizaje dado.
* Tienen una gran precisión de predicción.

Sin embargo, también cuenta con ciertos inconvenientes:

* El riesgo de sobreajuste crece a medida que el número de observaciones es más pequeño, lo que dará lugar a predicciones incorrectas.
* Si contamos con un gran número de variables, se consumirá un gran tiempo de computación.

**Embedded methods (métodos integrados)**: Combinan el método de filtrado y el de envoltura con su propio método de selección de características, y realizan la selección de características durante el entrenamiento del modelo.



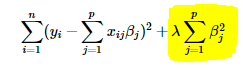
Vamos a profundizar en esta categoría de métodos viendo los principales tipos:

* Métodos basados en regularización: Añaden una penalización a los parámetros de un modelo dado con el fin de evitar sobreajustes. Hay 3 tipos:
  + Regresión de lazo: Añade una penalización a la función coste de la forma:



Siendo la sección señalada en amarillo la penalización referida.

* + Regresión de cresta: Al igual que la regresión de lazo, añade una penalización a la función coste, pero de forma cuadrática:



* Métodos basados en árboles: Tienen en cuenta la importancia de las características a la hora de seleccionarlas, de modo que obtenemos las variables con mayor importancia al ahora de dar un resultado fiable. Estos árboles de decisión están formados por un conjunto de características aleatorias del problema a tratar, de modo que un árbol no podrá acceder a las demás características. Cada nodo simboliza una condición de una característica, de modo que cada nodo divide el conjunto de datos en dos conjuntos distintos.

Usando este tipo de métodos se obtienen grandes beneficios, esto debido a combinar el uso de métodos de filtro y de envoltura, obteniendo así las ventajas de cada uno:

* Tienen en cuenta la interacción entre características.
* Son más rápidos y precisos que los métodos de envoltura.
* Evitan en gran medida el sobreajuste.
* Encuentra el subconjunto óptimo de características del problema a tratar.

Por consecuencia de lo anterior, presenta los siguientes inconvenientes:

* Implementación más difícil.
* Gran coste computacional.

Una vez vistos en profundidad cada una de las familias de métodos de selección de características, vamos a proceder a realizar una comparación entre los métodos de filtrado y los métodos de envoltura (no se contempla incluir los métodos integrados debido a que usa ambos métodos anteriores):

* Los métodos de envoltura entrenan con un subconjunto y miden su utilidad, mientras que los métodos de filtrado miden la correlación de las características con la variable dependiente.
* Los métodos de filtrado son mucho más rápidos debido a que no se entrena el modelo.
* Los métodos de envoltura suponen más coste computacional debido al entrenamiento del modelo del que carecen los métodos de filtrado.
* Los métodos de filtro evalúan grupos de características mediante métodos estadísticos; por el contrario, los métodos de envoltura los evalúan mediante validación cruzada.
* Los métodos de envoltura, a diferencia de los de filtrado, siempre encuentran el subconjunto óptimo de características.
* El subconjunto de características de los métodos de envoltura produce más riesgo de sobreajuste que los de los métodos de filtrado.
* Los métodos integrados son más rápidos que los métodos de envoltura, ya que al usar métodos de filtro simplifican el problema.
* Los métodos integrados suponen un mayor coste computacional respecto a los métodos de filtro y de envoltura, debido a que selecciona las características mientras se entrena el modelo a tratar.

En vista de estas observaciones, podemos dar como conclusión de este marco teórico los casos de uso de cada familia de métodos de filtrado en base al problema a tratar:

* Los métodos de filtrado serán de mayor utilidad cuando se nos presenta un conjunto con muchas características, de modo que nos interesará desechar aquellas que aporten poca información al modelo.
* Los métodos de envoltura nos serán de utilidad cuando queramos obtener el resultado más optimo en un conjunto de características de relativo tamaño, de modo que evitaríamos el sobreajuste al no ser este de un tamaño pequeño, y siempre y cuando no nos importe el coste computacional.
* Los métodos integrados, al igual que los de envoltura, nos devolverán la solución más óptima, aunque estos debido a su complejidad es más conveniente usarlos cuando tenemos un conjunto de características de gran tamaño, de modo que se comenzaría simplificando el problema usando la técnica de filtrado conveniente para posteriormente pasar a usar el método de envoltura deseado.

Bibliografía

<https://www.analyticsvidhya.com/blog/2016/12/introduction-to-feature-selection-methods-with-an-example-or-how-to-select-the-right-variables>

[https://es.wikipedia.org/wiki/Selecci%C3%B3n\_de\_variable#cite\_note-ReferenceA-19](https://es.wikipedia.org/wiki/Selecci%C3%B3n_de_variable%23cite_note-ReferenceA-19)

<https://es.wikipedia.org/wiki/Coeficiente_de_correlaci%C3%B3n_de_Pearson>

<https://en.wikipedia.org/wiki/Linear_discriminant_analysis>

<https://www.cienciadedatos.net/documentos/28_linear_discriminant_analysis_lda_y_quadratic_discriminant_analysis_qda>

<https://es.wikipedia.org/wiki/Distribuci%C3%B3n_%CF%87%C2%B2>

<https://heartbeat.fritz.ai/hands-on-with-feature-selection-techniques-wrapper-methods-5bb6d99b1274>

<https://heartbeat.fritz.ai/hands-on-with-feature-selection-techniques-embedded-methods-84747e814dab>

<https://www.datavedas.com/embedded-methods/>

<https://towardsdatascience.com/l1-and-l2-regularization-methods-ce25e7fc831c>

[https://www.analyticsvidhya.com/blog/2016/01/ridge-lasso-regression-python-complete-tutorial/#three](https://www.analyticsvidhya.com/blog/2016/01/ridge-lasso-regression-python-complete-tutorial/%23three)