**Drop Back Elimination**

Dữ liệu độ hấp thụ của mẫu ở trên dãy bước sóng được rời rạc hóa thành độ hấp thụ trên từng giá trị bước sóng xác định , đây chính là các đặc trưng trong bài toán phân loại. Ta thực hiện xác định hệ số tương quan giữa từng đặc trưng này đối với biến mục tiêu. Hệ số tương quan được tính toán qua công thức:

Trong đó:

* là độ hấp thụ của các mẫu ở bước sóng i
* là hàm lượng chất có trong mẫu
* là kỳ vọng

Tiếp theo, sắp xếp các bước sóng theo sự giảm dần của trị tuyệt đối hệ số tương quan. Sau đó lần lượt loại bỏ các bước sóng có độ tương quan thấp, sau mỗi lần loại bỏ áp dụng mô hình Stacked Regression kết hợp cross validation để kiểm tra. Việc kiểm tra được thực hiện như sau:

* Xác định giá trị của mô hình Stacked Regression đối với tập dữ liệu khi chưa loại bỏ các đặc trưng
* Xác định giá trị của mô hình Stacked Regression đối với tập dữ liệu khi loại bỏ i đặc trưng
* Tính toán độ mức ý nghĩa này qua công thức:

Trong đó: là độ lệch chuẩn của biến mục tiêu

* So sánh mức ý nghĩa này với ngưỡng α:
  + Nếu độ thì tiếp tục loại bỏ đặc trưng tiếp theo
  + Nếu độ thì dừng quá trình và lưu lại đặc trưng vừa loại bỏ
* Tập các đặc trưng còn lại chính là khoảng bước sóng mà chất hấp thụ mạnh nhất