Các phương pháp cải tiến Linear Regression

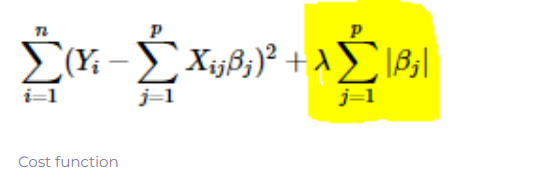
1. Regularization

**Regularization** trong linear regression là một kỹ thuật nhằm ngăn chặn hiện tượng overfitting (quá khớp) của mô hình hồi quy tuyến tính. Overfitting xảy ra khi mô hình học quá sát dữ liệu huấn luyện, dẫn đến khả năng tổng quát hóa kém trên dữ liệu mới. Regularization giải quyết vấn đề này bằng cách điều chỉnh hàm chi phí (cost function) để hạn chế độ phức tạp của mô hình và ưu tiên những mô hình đơn giản hơn.

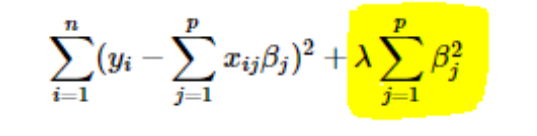
**Cách thức hoạt động:**

Thêm một số hạng mới vào hàm chi phí:

* L1 Regularization (Lasso Regression): Thêm tổng trị tuyệt đối của các tham số (theta).



* nếu lambda bằng 0 thì chúng ta sẽ nhận lại OLS (bình phương tối thiểu thông thường) trong khi một giá trị rất lớn sẽ làm cho các hệ số bằng 0, nghĩa là nó sẽ trở nên không phù hợp.
* L2 Regularization (Ridge Regression): Thêm tổng bình phương của các tham số (theta).



* nếu lambda bằng 0 thì trở lại lại OLS. Tuy nhiên, nếu lambda quá lớn thì nó sẽ tăng thêm trọng lượng và dẫn đến trang bị thiếu. Phải nói rằng, cách chúng ta chọn lambda rất quan trọng. Kỹ thuật này hoạt động rất tốt để tránh các vấn đề về overfitting.

**Sự khác biệt chính giữa các kỹ thuật này là lasso thu nhỏ hệ số của tính năng ít quan trọng hơn về 0, do đó loại bỏ hoàn toàn một số tính năng. Nói cách khác, chính quy hóa L1 hoạt động tốt cho việc lựa chọn tính năng trong trường hợp chúng ta có nhiều tính năng (features)**

**Ảnh hưởng:**

* Thu nhỏ các tham số (theta) về 0, khiến một số đặc trưng (feature) ít quan trọng hơn.
* Giảm độ phức tạp của mô hình, giúp giảm thiểu overfitting.

**Các loại Regularization thường gặp:**

* L1 Regularization: - Tốt cho việc chọn lọc đặc trưng (feature selection).Có thể đưa một số tham số về 0, loại bỏ chúng khỏi mô hình.
* L2 Regularization: - Không loại bỏ hoàn toàn các đặc trưng. Thường được sử dụng phổ biến hơn vì nó tạo ra các mô hình ổn định hơn.

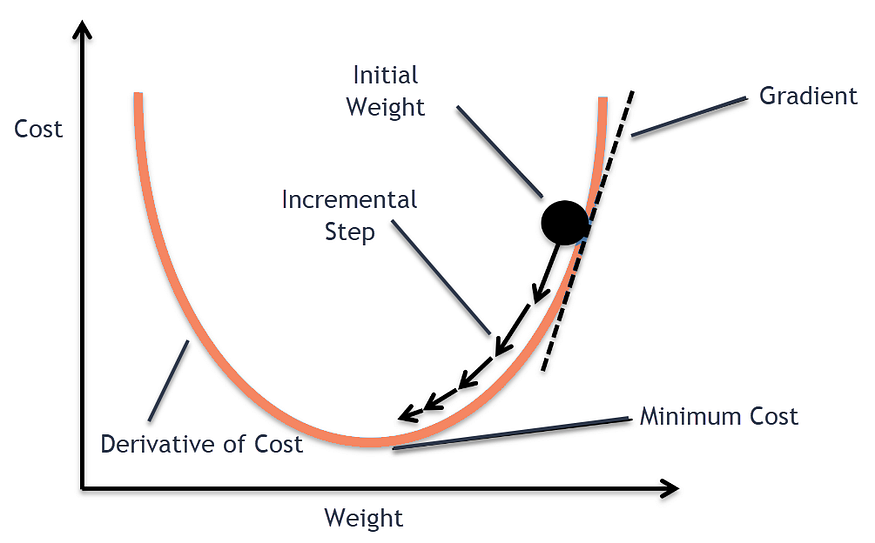
**Lựa chọn tham số lambda:**

* Tham số lambda kiểm soát mức độ ảnh hưởng của regularization.
* Lambda lớn hơn dẫn đến regularization mạnh hơn và mô hình đơn giản hơn.
* Lambda tối ưu thường được tìm kiếm thông qua kỹ thuật cross-validation.

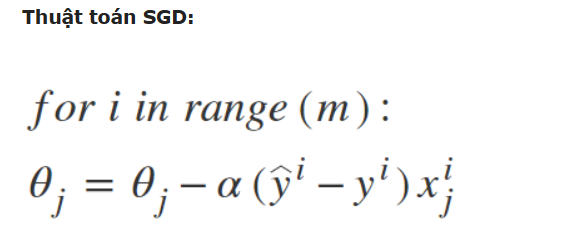
2. Gradient Descent Variants

- Gradient Descent Variants: Gradient Descent là một kỹ thuật tối ưu hóa phổ biến trong ML và Học sâu và nó có thể được sử dụng với hầu hết, các thuật toán học tập. Gradient là độ dốc của một hàm. Nó đo lường mức độ thay đổi của một biến để đáp ứng với những thay đổi của một biến khác. Về mặt toán học, Gradient Descent là một hàm lồi có đầu ra là đạo hàm riêng của một tập các tham số đầu vào của nó. Gradient càng lớn thì độ dốc càng lớn.

Bắt đầu từ một giá trị ban đầu, Gradient Descent được chạy lặp đi lặp lại để tìm các giá trị tối ưu của các tham số để tìm giá trị nhỏ nhất có thể có của chi phí hàm đã cho.



- SDG: Stochastic Gradient Descent (SGD) là một thuật toán tối ưu hóa được sử dụng trong học máy, đặc biệt là trong huấn luyện các mô hình mạng nơ-ron. Khác với Gradient Descent truyền thống, SGD chỉ sử dụng một mẫu duy nhất trong mỗi lần lặp để tính toán gradient. Điều này giúp SGD trở nên hiệu quả hơn về mặt tính toán, đặc biệt là với các tập dữ liệu lớn.



Ưu điểm của SGD:

* Hiệu quả về mặt tính toán
* Có thể thoát khỏi các điểm cực tiểu cục bộ
* Dễ dàng thực hiện

Nhược điểm của SGD:

* Độ ổn định có thể kém hơn so với Gradient Descent truyền thống
* Cần điều chỉnh cẩn thận các tham số học tập

Batch Gradient Descent

* Là một thuật toán tối ưu hóa được sử dụng để tìm điểm tối thiểu của một hàm mất mát.
* Để làm vậy, Batch Gradient Descent sử dụng tất cả các điểm dữ liệu trong tập huấn luyện để tính toán đạo hàm của hàm mất mát tại mỗi lần lặp.

Ưu điểm:

* Độ ổn định cao
* Có thể đạt được độ chính xác cao

Nhược điểm:

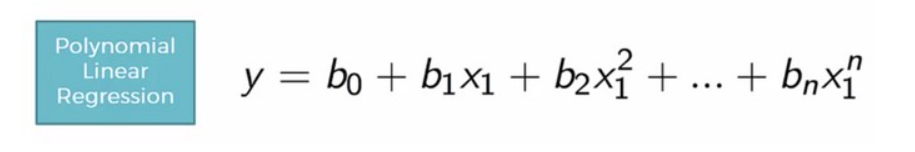
* Không hiệu quả với các tập dữ liệu lớn
* Không hiệu quả với online learning

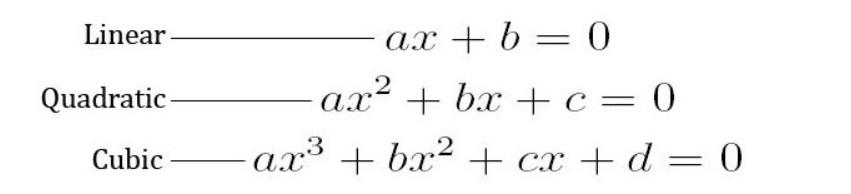
3. Polynomial Regression

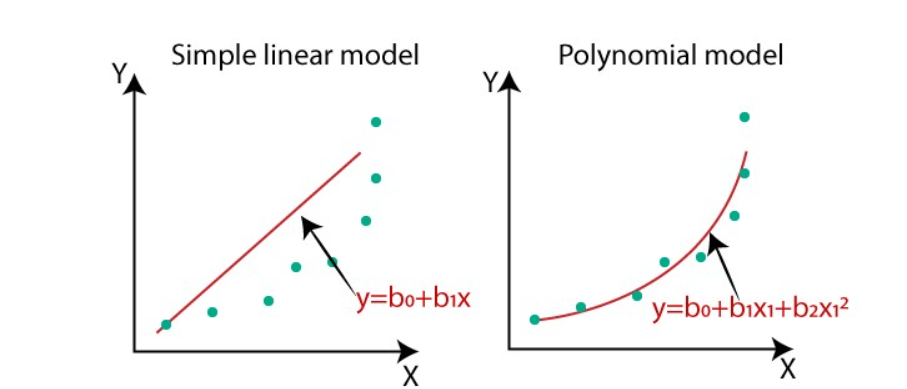
Polynomial Regression là gì?

Polynomial Regression là thuật toán hồi quy đa thức, nó giống như thuật toán hồi quy tuyến tính, sử dụng mối quan hệ giữa các biến độc lập x và biến phụ thuộc y được biểu diễn dưới dạng đa thức bậc n, để tìm cách tốt nhất vẽ một đường qua các điểm dữ liệu sao cho tối ưu và phù hợp nhất. Polynomial Regression là một thuật toán trong machine learning, nó được dùng cho các bài toán về dự đoán, dự báo (prediction).

Dạng tổng quát của Polynomial Regression như sau:



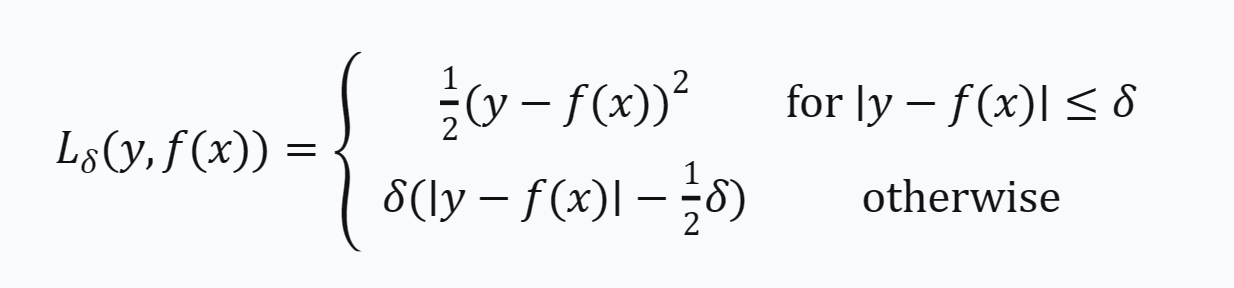
Một số dạng thường gặp của Polynomial Regression: 



4. Huber Loss

Huber Loss là một hàm mất mát được sử dụng trong linear regression để giảm ảnh hưởng của các điểm ngoại lệ (outliers) so với Mean Squared Error (MSE). Trong mô hình linear regression, MSE có thể là nhạy cảm với các giá trị ngoại lệ và dẫn đến việc mô hình hóa không chính xác.

Huber Loss được định nghĩa như sau:



Trong đó:

* y là giá trị thực tế
* f(x) là giá trị dự đoán
* δ là một hằng số dương.
* Khi ∣y−f(x)∣≤δ, hàm loss sử dụng bình phương sai số giữa giá trị dự đoán và giá trị thực tế.
* Khi ∣y−f(x)∣>δ, hàm loss sử dụng hàm tuyến tính để giảm thiểu ảnh hưởng của các giá trị ngoại lai.