Робаста регресія

Дуже часто зустрічається моделювання шуму в регресійній моделі за допомогою гауссового розподілу. Однак, якщо в наших даних є outliers, що в результати мають їх, це може призвести до поганої форми. Це тому, що помилка в квадраті карається відхиленням квадратично, тому точка, що знаходиться далеко від лінії, більше впливає на прилягання, ніж точки, розташовані поблизу лінії.

Одним із способів досягти стійкості для outliers є заміна розподілу Гаусса для змінної реакції на розподіл, який має важкі хвости, наприклад розподіл Лапласа. Тоді ми отримуємо ймовірність складання:



Один із способів обчислити ваговий вектор моделі за ймовірністю Лапласа - це мінімізувати функцію втрат Хубера (Huber, 1964):

Альтернативний спосіб полягає в мінімізації негативної ймовірності:



На жаль, це нелінійна цільова функція, яку важко оптимізувати. На щастя, ми можемо перетворити NLL в лінійну ціль, залежно від лінійних обмежень, використовуючи трюк роздільної змінної:

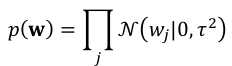


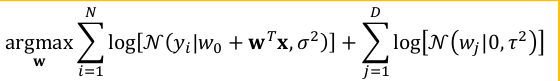


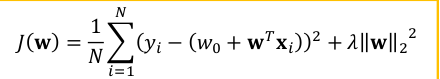
Ridge regression

Однією з проблем оцінки ML є те, що це може призвести до перевиконання - якби ми трохи змінили дані, отримані коефіцієнти сильно змінилися.

Ми можемо заохотити параметри MLE бути невеликими, таким чином, в результаті вийде більш плавна крива, використовуючи нульове середнє значення Гаусса:







Ця техніка називається найменшими штрафами. Загалом, додавання Гаусса до параметрів моделі, щоб заохотити їх бути невеликими, називається регуляризацією l 2 або занепадом ваги.

