Функция потерь:

Где i – номер наблюдения, xi, yi – фиксированные (константные) значения.

f(xi,yi) – выходное константное значение.

Производная функции потерь по переменной a:

Аналогично переменной b получаем:

Получили матрицу с коэффициентами:

|  |  |  |
| --- | --- | --- |
| Переменная «a» | Переменная «b» | Свободный член |
|  |  |  |
|  |  |  |

Важно понимать, что это ситуация может масштабироваться. Вместо переменной «a», «b» или любой другой может быть любая функция от этой переменной f(x) или даже функция от нескольких переменных F(a, b). Так как любое F(ai, bi, ci …) – это просто константное значение диктуемое наблюдаемыми значениями и вся суть МНК сводиться к определению значений переменных функции потерь, таких, что при данных константных значениях, задаваемыми прямыми или обработанными наблюдениями значение функции потерь минимально.

Функция является параболоидом. Так как, если для каждого i наблюдения выполнить для внутреннего содержимого возведение в степень получится уравнение вида k1 \* a2 + k2 \* b2 + k3 \* a \* b + Const.

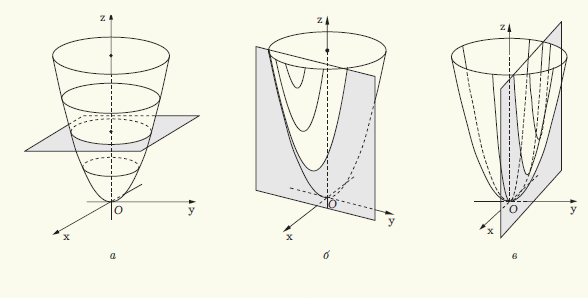
При раскрытии степени вынесем за скобки a, b =>

,

где и это сумма фиксированных (константных) наблюдений.

Получаем примерную зависимость вида , что есть параболоид.

Легко понять, если представить, что «a» это переменная, а «b» константа, то мы находимся и можем двигаться по двумерному срезу такого параболоида.



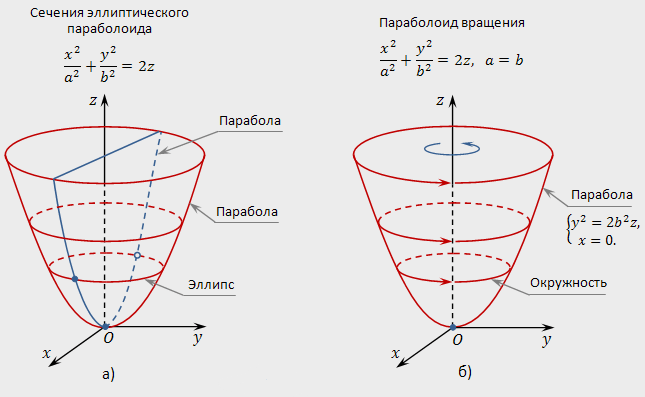


Рисунок 3 – параболоид

В нашем случае речь идет об эллиптическом параболоиде. Коэффициенты k1 и k2 определяют "растянутость" эллиптического параболоида вдоль осей x и y соответственно.

… ,

Так как, например k1 представимо в виде , где C1 некая величина.

Важность наблюдения заключается в том, что у параболоида вида производная по любой переменной растущая, а не убывающая. Можно мысленно представить, что находя в каждом срезе минимум и переходя к другому, будет отыскано некое минимальное значение функции, что гарантирует наличие одного глобального минимума.

Парабола может быть и большей мерности.

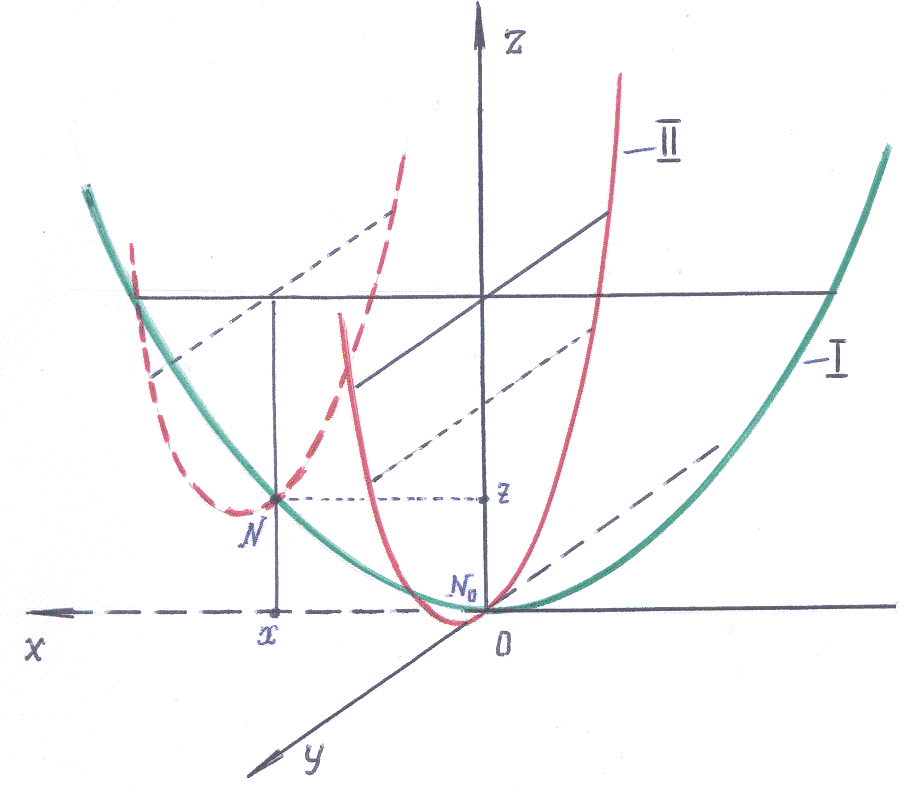


Рисунок 5 – параболоид

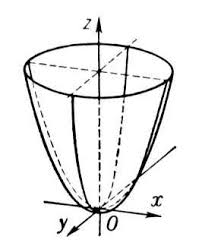


Рисунок 1 – параболоид

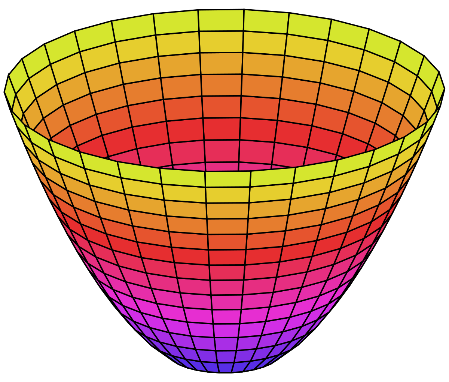


Рисунок 2 – параболоид

Регуляризация L1 и L2 заключается в дополнительном «оштрафовывании» функции потерь за слишком большой вес.

Функция потерь с L1 регуляризацией:

Если искать матрицу по предыдущему алгоритму, получиться следующее:

Для добавления L1 достаточно вычесть коэффициент из свободных членов, так как неважно в случае с L1 по какой переменной брать производную.

Функция потерь с L2 регуляризацией:

Разница в степени, для L2 квадрат суммы весов на коэффициент регуляризации. Очевидно, что здесь от выбора переменной, что-то будет зависеть. Если точнее от в зависимости от выбранной переменной мы будем добавлять + 2 \* L2 к коэффициенту перед этой переменной. Что значит, что после формирования матрицы, для применения такой регуляризации нужно будет проитерироваться по диагонали матрицы A.

Функция потерь с ElasticNet (L1 + L2) регуляризацией:

Очевидно, что для совмещенного варианта:

Для чего нужна регуляризация?

Слишком большие веса приводят к тому, что выходное значение функции может сильно меняться по мере роста переменных, малые веса дают сравнительно более «плавную картину».