

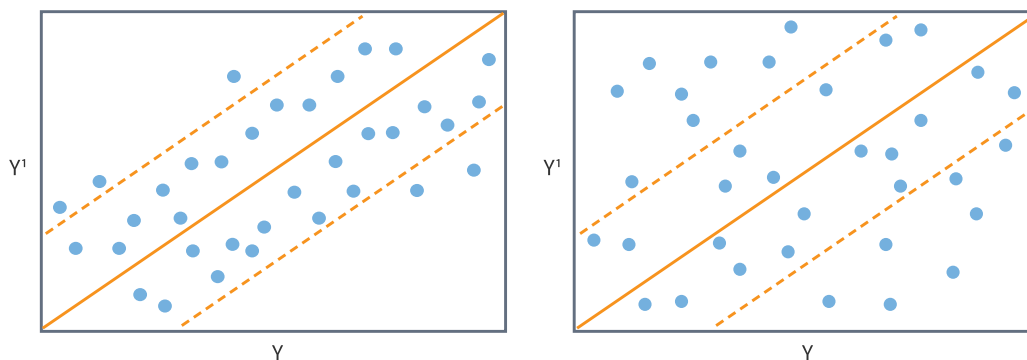
Диаграмма рассеяния (Scatter diagram)

Разделы: [Визуализация](#)

Диаграмма рассеяния — метод визуализации для оценивания точности работы регрессионных моделей. Представляет собой график, на котором по горизонтальной оси откладываются целевые (фактически наблюдаемые) значения обучающих примеров, а по вертикальной — значения, оцененные моделью. Тогда каждая пара целевого Y и оцененного Y' значений может быть представлена на диаграмме в виде точки.

Множество точек, для которых оцененное значение будет равно реальному, образует так называемую **линию идеальных значений**, для каждой точки которой выполняется равенство $Y = Y'$. Остальные точки, образованные примерами, на которых модель допустила ошибку, будут разбросаны относительно этой линии.

По степени их разброса можно оценить точность модели. Если большая часть точек сосредоточена вдоль идеальной линии, а значительные отклонения редки или совсем отсутствуют, то модель работает хорошо. Если разброс точек велик, то ее точность низкая.



На практике при визуализации диаграммы обычно строят две вспомогательные линии, соответствующие допустимому уровню ошибки $E = Y - Y'$ (пунктирные линии на рисунке). Если большая часть точек лежит в пределах установленного ими «коридора», то можно утверждать, что модель работает хорошо (левый рисунок). Если точки имеют значительный разброс и большая их часть расположена вне «коридора», то точность модели низкая (правый рисунок).

Допустимая ошибка определяется аналитиком, исходя из особенностей решаемой задачи.

