Метод стохастического градиента

1. В модели восстановления регрессии решение ищется в виде линейной функции: суммы признаков с коэффициентами (взвешенная сумма всех признаков):  
   Или с дополнительным коэффициентом :  
   Дополнительный коэффициент влияет на то, будет проходить функция через начало координат или нет.
2. Функция потерь в методе наименьших квадратов имеет следующий вид:
3. Минимум функции потерь в методе наименьших квадратов ищется путём дифференцирования функции по и условия равенства нулю полученного выражения (условие экстремума).
4. В линейной модели классификации решение ищется в виде функции знака числа (sign) от суммы всех признаков с коэффициентами :  
   Если – это направляющий вектор к разделяющей 2 класса гиперплоскости, то в зависимости от знака скалярного произведения точка попадает в один из этих классов.
5. С помощью пороговой функции потерь можно посчитать количество неправильных ответов, но она не говорит о величине ошибки – её нельзя продифференцировать и приравняв к нулю найти минимум.
6. Выражение отступа для объекта имеет вид:  
   При отсутствии ошибок пороговой функции потерь отступ получается положительным, а при присутствии – отрицательным. По отступу также можно судить о расстоянии точки от разделяющей гиперплоскости: чем больше отступ по модулю, тем дальше точка находится от гиперплоскости.
7. Кроме квадратичной функции потерь также существуют следующие функции потерь: средняя абсолютная ошибка (MSE), категориальная кросс-энтропия (CCE), бинарная кросс-энтропия (BCE) и другие.

Метод стохастического градиента

1. В методе стохастического градиента используется ускорение сходимости: вместо вычисления градиента на каждой итерации берётся по одной случайной точке и сразу обновляется вектор весов .
2. Из-за линейной зависимости признаков у задачи может быть не одно оптимальное решение, а целое семейство решений. В связи с этим метод становится неустойчивым и может сходиться к любой из этих точек, и на его ответ могут влиять отдельные точки обучающей выборки. Возникает переобучение, признаком которого являются большие по модулю веса разных знаков. Для борьбы с увеличением весов применяют регуляризацию.
3. В методе стохастического градиента задаются следующие параметры:
   1. loss = ‘hinge’ – функция потерь, по умолчанию метод опорных векторов.
   2. penalty = ‘l2’ – регуляризация суммой квадратов.
   3. alpha = 0.0001 – коэффициент регуляризации.
   4. l1\_ratio – коэффициент регуляризации, если penalty = ‘elasticnet’.
   5. learning\_rate = ‘optimal’ – шаг в формуле стохастического градиента. По умолчанию уменьшается по заданной формуле.
   6. eta = 0.0 – начальный шаг в формуле стохастического градиента.
4. Преимущества:
   1. Легко реализуется.
   2. Легко обобщается на нелинейные модели и любые функции потерь .
   3. Допускает потоковое обучение, когда не все данные загружаются.
   4. Подходит для больших данных – даёт неплохое решение, даже не обработав всю выборку.

Недостатки:

1. Возможна расходимость или медленная сходимость.
2. Застревание в локальных минимумах.
3. Много параметров, которые нужно настраивать.
4. Возможно переобучение из-за линейной зависимости признаков.