**Метод ближайших соседей** — простейший метрический классификатор, основанный на оценивании сходства объектов. Классифицируемый объект относится к тому классу, которому принадлежат ближайшие к нему объекты обучающей выборки.

**Метод k ближайших соседей**. Для повышения надёжности классификации объект относится к тому классу, которому принадлежит большинство из его *соседей* — ближайших к нему объектов обучающей выборки . В задачах с двумя классами число соседей берут нечётным, чтобы не возникало ситуаций неоднозначности, когда одинаковое число соседей принадлежат разным классам.

В наиболее общем виде алгоритм ближайших соседей:



где - i-ый сосед объекта u.

 - весовая функция.

1)  - метод k ближайших соседей.

Путем скользящего контроля подберем оптимальное значение параметра k = 11. Различие в качестве невелико для k< 16.

k=11,

'Accuracy of the procedure is ': '98.74%',

'Fall-out (FPR) of the procedure is ': '3.62%',

'False omission rate (FOR) of the procedure is ': '0.0%',

'Precision (PPV) of the procedure is ': '98.13%',

'Sensitivity (TPR) of the procedure is ': '100.0%'

k=20,

'Accuracy of the procedure is ': '79.62%',

'Fall-out (FPR) of the procedure is ': '58.6%',

'False omission rate (FOR) of the procedure is ': '0.83%',

'Precision (PPV) of the procedure is ': '76.43%',

'Sensitivity (TPR) of the procedure is ': '99.84%'}

2) - экспоненциальный вес(в данном случае q=0.8)

Как видно, использование экспоненциального веса незначительно, но улучшает работу алгоритма.

k=20, экспоненциальный вес

'Accuracy of the procedure is ': '80.7%',

'Fall-out (FPR) of the procedure is ': '55.37%',

'False omission rate (FOR) of the procedure is ': '0.77%',

'Precision (PPV) of the procedure is ': '77.4%',

'Sensitivity (TPR) of the procedure is ': '99.84%'}

k=11, экспоненциальный вес

'Accuracy of the procedure is ': '98.96%',

'Fall-out (FPR) of the procedure is ': '2.98%',

'False omission rate (FOR) of the procedure is ': '0.0%',

'Precision (PPV) of the procedure is ': '98.43%',

'Sensitivity (TPR) of the procedure is ': '100.0%'

3) Рассмотрим другую вариацию алгоритма ближайшего соседа. Зададим неоднородные веса клеткам поля "крестики-нолики". Так, например, центр участвует в 4 выигрышных комбинациях, верхний левый - в 3. Введем метрику - если в клетке стоит крестик или нолик, то считаем ее "важность" по количеству выигрышных комбинаций. Если пропуск, то - 1

'Accuracy of the procedure is ': '98.96%',

'Fall-out (FPR) of the procedure is ': '2.98%',

'False omission rate (FOR) of the procedure is ': '0.0%',

'Precision (PPV) of the procedure is ': '98.42%',

'Sensitivity (TPR) of the procedure is ': '100.0%'

Как видно, точность не отличается от метода ближайших соседей с экспоненциальным весом. Зато метод проигрывает в быстродействии.

4) Используем несколько другой подход - метод потенциальных функций.



здесь  - ядро.

h - потенциальное окно(выбирается эмпирически).

 - "важность" элемента. высчитывается по выборке или же принимаетсяна основании данных.

Возьмем метрику невзвешенную по различающимся признакам.

Результат:

'Accuracy of the procedure is ': '68.51%',

'Fall-out (FPR) of the procedure is ': '46.22%',

'False omission rate (FOR) of the procedure is ': '45.35%',

'Precision (PPV) of the procedure is ': '75.77%',

'Sensitivity (TPR) of the procedure is ': '76.26%'

Алгоритм плохо работает в силу произвольности важности( сделали центральную клетку сильно различающей объекты) - следует обучать коэффициенты. Таким образом слишком хорошие результаты скорее объясняются "хорошестью" набора для универсального метода ближайших соседей.