5. Метод k ближайших соседей в задачах классификации и восстановление регрессии. Теорема об оценке риска в методе ближайшего соседа. Идея доказательства.

Метод ближайшего соседа - 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 - k ближайших к нему объектов обучающей выборки x_i.

Пусть задана обучающая выборка: $\{(x^{(1)}, y^{(1)}), ..., (x^{(N)}, y^{(N)})\}$

$x_i \in X$ - Пространство признаков

$y \in Y$ - Выходы (responses)

~~Пусть на множестве объектов~~

$$f(x) = \frac{1}{k} \sum_{x^{(i)} \in N_k(X)} y^{(i)} \qquad f(x) = \arg\max_{y} |I_k(X,Y)|$$

где, $N_k(X)$ - множество из k ближайших к x (по евклидову расстоянию) точек из обучающей выборки. $I_k(X,Y)$ - множество тех точек $x^{(i)}$ из $N_k(X)$, для которых $y^{(i)} = y$. ~~$f: X \to Y$ - функция, которая проецирует входные данные из пространства~~ ~~X в~~ $f: X \to Y$ - функция, которые по пространству X предсказывает объекты из пространства Y.