Теорема об оценке риска (Cover, Hart, 1967):

Пусть R^* - оптимальное (байесовское) значение среднего риска для некоторой задачи классификации на K классов. Тогда с ростом размера выборки N ожидаемый риск R для метода одного ближайшего соседа сходится к R^0, такому, что

$$R^* \le R^0 \le R^* \cdot \left(2 - \frac{K}{K-1} R^*\right) \le 2R^*$$

Доказательство:

x' - объект из обучающей выборки, ближайший к x, а Y' - соответствующий выход

$k^* = \arg\max Pr(k|x)$ - байесово решение (самый популярный класс) в точке x

$r^*(x) = 1 - Pr(k^*|x)$ предельный условный байесовский риск

$r(x)$ - предельный условный средний риск классификатора ближайшего соседа

$$r(x) = \lim_{N\to\infty} Pr(Y \ne Y' | x) = \lim_{N\to\infty} \int_X Pr(Y = Y' | x, x') p(x'|x) dx' =$$

$$= \lim_{N\to\infty} \int_X \left(1 - \sum_{k=1}^{K} Pr(k|x) Pr(k|x')\right) p(x'|x) dx' = \int_X \left(1 - \sum_{k=1}^{K} Pr(k|x) Pr(k|x')\right) \delta(x'-x) dx' =$$

$$= 1 - \sum_{k=1}^{K} Pr^2(k|x) = 1 - Pr^2(k^*|x) - \sum_{k \ne k^*} Pr^2(k|x) \le 1 - Pr^2(k^*|x) - \frac{1}{K-1}\left(\sum_{k \ne k^*} Pr(k|x)\right)^2 =$$

$$= 1 - Pr^2(k^*|x) - \frac{1}{K-1}\left(1 - Pr(k^*|x)\right)^2 = 1 - (1 - r^*(x))^2 - \frac{1}{K-1} r^*(x) = 2r^*(x) - \frac{K}{K-1} r^*(x)^2$$

где, R - ожидаемая ошибка предсказания $R(f) = EL(f(X), Y) = \int_{X \times Y} L(f(x), y) p(x,y) dx dy$

L - функция потерь

квадратичная $L(y', y) = (y' - y)^2$

абсолютная $L(y', y) = |y' - y|$

относительная $L(y', y) = \frac{|y' - y|}{|y|}$

симметричная $L(y', y) = I(y' \ne y) \equiv \begin{cases} 0, & y' = y \\ 1, & y' \ne y \end{cases}$