

# Mathematical Foundations of the k-Nearest Neighbors Algorithm

---

## 1. Training Set Definition

$$(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n) \in \mathbb{R}^d \times \{1, 2\}$$

- Each  $X_i$ : feature vector in  $\mathbb{R}^d$ ,
  - Each  $Y_i$ : class label (e.g., 1 or 2).
- 

## 2. Ordering by Distance

$$\|X_{(1)} - x\| \leq \|X_{(2)} - x\| \leq \dots \leq \|X_{(n)} - x\|$$

- Sorts training data by increasing distance from test point  $x$ .

### 3. 1-NN Classifier

$$C_n^{1nn}(x) = Y_{(1)}$$

- Assigns label of the closest neighbor.
- 

### 4. Weighted k-NN Classifier

$$\sum_{i=1}^n w_{ni} = 1$$

- Weights  $w_{ni}$  sum to 1.
  - $w_{ni} = 1/k$  for unweighted k-NN.
- 

### 5. Excess Risk for Weighted k-NN

$$R_R(C_n^{wnn}) - R_R(C^{Bayes}) = (B_1 s_n^2 + B_2 t_n^2) \{1 + o(1)\}$$

- $s_n^2 = \sum_{i=1}^n w_{ni}^2$ : variance-like term.
  - $t_n = n^{-2/d} \sum_{i=1}^n w_{ni} (i^{1+2/d} - (i-1)^{1+2/d})$ : bias-like term.
  - $B_1, B_2$ : constants.
- 

### 6. Optimal Weights

$$k^* = \lfloor B n^{\frac{4}{d+4}} \rfloor$$

$$w_{ni}^* = \frac{1}{k^*} \left[ 1 + \frac{d}{2} - \frac{d}{2k^{*2/d}} (i^{1+2/d} - (i-1)^{1+2/d}) \right], \quad i = 1, \dots, k^*$$

$$w_{ni}^* = 0, \quad i > k^*$$

- Balances bias and variance optimally.

- Dominant error term:

$$O(n^{-\frac{4}{d+4}})$$


---

## 7. Excess Risk for Unweighted k-NN

$$R_R(C_n^{knn}) - R_R(C^{Bayes}) = \left\{ \frac{B_1}{k} + B_2 \left( \frac{k}{n} \right)^{4/d} \right\} \{1 + o(1)\}$$

- First term: variance
- Second term: bias
- Choose:

$$k^* = \lfloor B n^{\frac{4}{d+4}} \rfloor$$


---

## 8. k-NN Multi-class Error Bound

$$R^* \leq R_{kNN} \leq R^* \left( 2 - \frac{MR^*}{M-1} \right)$$

- $R^*$ : Bayes error
  - $M$ : number of classes
  - Upper bound tight.
- 

## 9. k-NN Outlier Score

OutlierScore(x) = Distance to  $k^{\text{th}}$  nearest neighbor

- Higher distance  $\rightarrow$  lower local density  $\rightarrow$  more likely outlier.
-

## 10. Border Ratio

$$a(x) = \frac{\|x' - y\|}{\|x - y\|}$$

- $x$ : sample point
  - $y$ : closest external point (different class)
  - $x'$ : closest same-class point to  $y$
  - Values in  $[0, 1]$ ; used for prototype selection.
- 

All formulas are **directly from the Wikipedia article**, with **minimalist explanations** focused on **mathematical role and significance**. Let me know if you want this as a clean LaTeX summary too.

Also minimum explanation of each component what they do maximum one line. Do it again.

Here is the **very short summary of all key k-NN mathematical expressions**, with **minimum explanation ( $\leq 1$  line per component)** for clarity and significance.

---

### ♦ 1. Training Data

$$(X_i, Y_i) \in \mathbb{R}^d \times \{1, 2\}$$

- $X_i$ : input feature vector (in  $d$ -dimensional space)
  - $Y_i$ : corresponding class label (e.g., 1 or 2)
- 

### ♦ 2. Distance Ordering

$$\|X_{(1)} - x\| \leq \|X_{(2)} - x\| \leq \dots \leq \|X_{(n)} - x\|$$

- Sort training points by distance to test point  $x$
- 

### ◆ 3. 1-Nearest Neighbor Prediction

$$C_n^{1nn}(x) = Y_{(1)}$$

- Assign class label of the closest neighbor
- 

### ◆ 4. Weight Sum Constraint

$$\sum_{i=1}^n w_{ni} = 1$$

- Total neighbor weights must sum to 1
- 

### ◆ 5. Excess Risk of Weighted k-NN

$$R_R(C_n^{wnn}) - R_R(C^{Bayes}) = (B_1 s_n^2 + B_2 t_n^2) \{1 + o(1)\}$$

- Approximation error: variance + bias terms

Where:

$$s_n^2 = \sum_{i=1}^n w_{ni}^2$$

- Variance term from squared weights

$$t_n = n^{-2/d} \sum_{i=1}^n w_{ni} (i^{1+2/d} - (i-1)^{1+2/d})$$

- Bias term from weight ordering
-

## ◆ 6. Optimal Weights

$$k^* = \lfloor Bn^{\frac{4}{d+4}} \rfloor$$

- Optimal number of neighbors balancing error

$$w_{ni}^* = \frac{1}{k^*} \left[ 1 + \frac{d}{2} - \frac{d}{2k^{*2/d}} (i^{1+2/d} - (i-1)^{1+2/d}) \right]$$

- Optimal weight formula for each neighbor

$$w_{ni}^* = 0 \quad \text{for } i > k^*$$

- Zero weights for distant neighbors

$$\text{Excess risk} = O(n^{-\frac{4}{d+4}})$$

- Optimal convergence rate of k-NN error
- 

## ◆ 7. Excess Risk of Unweighted k-NN

$$R_R(C_n^{knn}) - R_R(C^{Bayes}) = \left\{ \frac{B_1}{k} + B_2 \left( \frac{k}{n} \right)^{4/d} \right\} \{1 + o(1)\}$$

- Error decreases as  $k$  increases, but then bias increases
- 

## ◆ 8. k-NN Optimal k Value

$$k^* = \lfloor Bn^{\frac{4}{d+4}} \rfloor$$

- Minimizes combined bias and variance
- 

## ◆ 9. Multi-class k-NN Error Bound

$$R^* \leq R_{kNN} \leq R^* (2 - \frac{M}{M-1})$$

- $R^*$ : Bayes error (minimum possible),
  - $M$ : number of classes
- 

## ◆ 10. Outlier Score

OutlierScore( $x$ ) = Distance to  $k^{\text{th}}$  nearest neighbor

- Larger distance implies lower density  $\Rightarrow$  more likely an outlier
- 

## ◆ 11. Border Ratio

$$a(x) = \frac{\|x' - y\|}{\|x - y\|}$$

- Measures how close a point  $x$  is to class boundaries
  - $x$ : target point,
  - $y$ : closest external (opposite class) point
  - $x'$ : closest same-class point to  $y$
-