T4: Nonparametric and distance-based learning

Fundamentos del Aprendizaje Automático

Curso 2025/2026

Structure

Outline

Bayesian statistical learning

$$P(\omega|\mathbf{x}) = \frac{p(\mathbf{x}|\omega) \cdot P(\omega)}{p(\mathbf{x})}$$

Bayesian statistical learning

$$P(\omega|\mathbf{x}) = \frac{p(\mathbf{x}|\omega) \cdot P(\omega)}{p(\mathbf{x})}$$

- **Ideal** case: $p(\mathbf{x}|\omega)$ and $P(\omega)$ are known and well defined

Bayesian statistical learning

$$P(\omega|\mathbf{x}) = \frac{p(\mathbf{x}|\omega) \cdot P(\omega)}{p(\mathbf{x})}$$

- **Ideal** case: $p(\mathbf{x}|\omega)$ and $P(\omega)$ are known and well defined
- **Practical** case: $p(\mathbf{x}|\omega)$ and $P(\omega)$ are estimated from assortment $\mathcal{D}$

Bayesian statistical learning

$$P(\omega|\mathbf{x}) = \frac{p(\mathbf{x}|\omega) \cdot P(\omega)}{p(\mathbf{x})}$$

- **Ideal** case: $p(\mathbf{x}|\omega)$ and $P(\omega)$ are known and well defined
- **Practical** case: $p(\mathbf{x}|\omega)$ and $P(\omega)$ are estimated from assortment $\mathcal{D}$

How is it estimated?

Approach	**Module**	**Prior** $P(\omega)$	**Likelihood** $p(\mathbf{x}|\omega)$	
			Assumption	**Estimation**
Parametric	T2	Frequentist	Multivariate distribution Statistical independence and univariate distributions	Maximum Likelihood Estimation
Nonparametric	T4	Frequentist	No distribution is assumed	Histogram, Parzen windows, Nearest Neighbor

Nonparametric learning

Two scenarios:

Nonparametric learning

Two scenarios:

1. Estimate *likelihoods* $p(\mathbf{x}|\omega)$ and *priors* $P(\omega)$

Nonparametric learning

Two scenarios:

1. Estimate *likelihoods* $p(\mathbf{x}|\omega)$ and *priors* $P(\omega)$

2. Estimate *posteriors* $P(\omega|\mathbf{x})$

Nonparametric learning

Two scenarios:

1. Estimate *likelihoods* $p(\mathbf{x}|\omega)$ and *priors* $P(\omega)$
   - **Key procedure**: (probability) density estimation

2. Estimate *posteriors* $P(\omega|\mathbf{x})$

Nonparametric learning

Two scenarios:

1. Estimate *likelihoods* $p(\mathbf{x}|\omega)$ and *priors* $P(\omega)$
   - **Key procedure**: (probability) density estimation
     - $\rightarrow$ Reconstructing the probability density function of a random variable
     - $\rightarrow$ Independent of the true distribution

2. Estimate *posteriors* $P(\omega|\mathbf{x})$

Nonparametric learning

Two scenarios:

1. Estimate *likelihoods* $p(\mathbf{x}|\omega)$ and *priors* $P(\omega)$
   - **Key procedure**: (probability) density estimation
     - $\rightarrow$ Reconstructing the probability density function of a random variable
     - $\rightarrow$ Independent of the true distribution
   - The function approximates the true likelihood $\Rightarrow \hat{p}(\mathbf{x}|\omega)$
2. Estimate *posteriors* $P(\omega|\mathbf{x})$

Nonparametric learning

Two scenarios:

1. Estimate *likelihoods* $p(\mathbf{x}|\omega)$ and *priors* $P(\omega)$
   - **Key procedure**: (probability) density estimation
     - $\rightarrow$ Reconstructing the probability density function of a random variable
     - $\rightarrow$ Independent of the true distribution
   - The function approximates the true likelihood $\Rightarrow \hat{p}(\mathbf{x}|\omega)$
2. Estimate *posteriors* $P(\omega|\mathbf{x})$
   - Avoids *likelihood* and *prior* estimation

Outline

Formulation

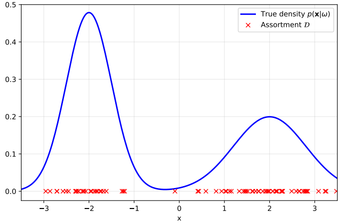
- Data assortment $\mathcal{D} = \{(\mathbf{x}_i, \omega_i)\}_{i=1}^{|\mathcal{D}|} \subset \mathbb{R}^d \times \mathcal{W}$

Formulation

- Data assortment $\mathcal{D} = \{(\mathbf{x}_i, \omega_i)\}_{i=1}^{|\mathcal{D}|} \subset \mathbb{R}^d \times \mathcal{W}$
  - $\rightarrow$ **Goal**: estimate the unknown $p(\mathbf{x}|\omega)$ for each label

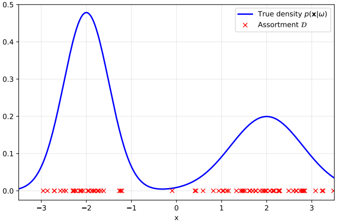
Formulation

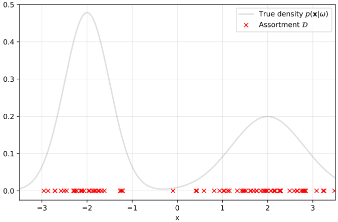
- Data assortment $\mathcal{D} = \{(\mathbf{x}_i, \omega_i)\}_{i=1}^{|\mathcal{D}|} \subset \mathbb{R}^d \times \mathcal{W}$
  - $\rightarrow$ **Goal**: estimate the unknown $p(\mathbf{x}|\omega)$ for each label



Formulation

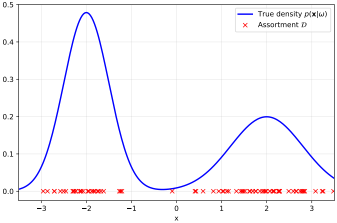
- Data assortment $\mathcal{D} = \{(\mathbf{x}_i, \omega_i)\}_{i=1}^{|\mathcal{D}|} \subset \mathbb{R}^d \times \mathcal{W}$
  $\rightarrow$ **Goal**: estimate the unknown $p(\mathbf{x}|\omega)$ for each label

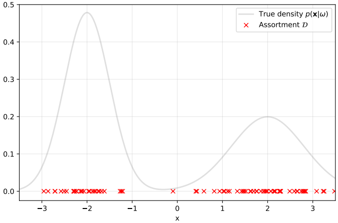




Formulation

- Data assortment $\mathcal{D} = \{(\mathbf{x}_i, \omega_i)\}_{i=1}^{|\mathcal{D}|} \subset \mathbb{R}^d \times \mathcal{W}$
  - $\rightarrow$ **Goal**: estimate the unknown $p(\mathbf{x}|\omega)$ for each label

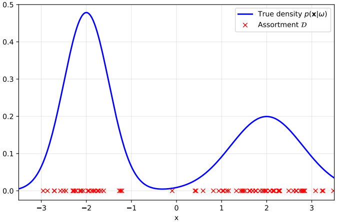


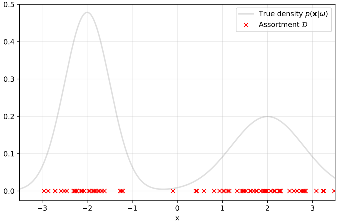


- Histogram estimator:
  - $\rightarrow$ Model $p(\mathbf{x}|\omega)$ as a histogram
  - $\rightarrow$ **Process**:
    1. Divide data space into regions
    2. Count samples per region
    3. Density as a ratio

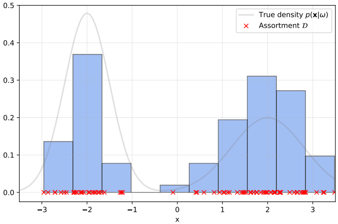
Formulation

- Data assortment $\mathcal{D} = \{(\mathbf{x}_i, \omega_i)\}_{i=1}^{|\mathcal{D}|} \subset \mathbb{R}^d \times \mathcal{W}$
  - $\rightarrow$ **Goal**: estimate the unknown $p(\mathbf{x}|\omega)$ for each label





- Histogram estimator:
  - $\rightarrow$ Model $p(\mathbf{x}|\omega)$ as a histogram
  - $\rightarrow$ **Process**:
    1. Divide data space into regions
    2. Count samples per region
    3. Density as a ratio



Formulation

- Assume a set of regions $\{\mathcal{R}_1, \ldots \mathcal{R}_N\}$ that divide the data space

Formulation

- Assume a set of regions $\{\mathcal{R}_1, \ldots \mathcal{R}_N\}$ that divide the data space
- The probability of $\mathbf{x} \in \mathcal{D}_\omega$ being in region $\mathcal{R}_j$:

$$P(\mathbf{x} \in \mathcal{R}_j) = \int_{\mathcal{R}_j} p(\mathbf{u})\, d\mathbf{u} \approx p(\mathbf{x}|\omega) \cdot V \text{ where } V \in \mathbb{R}^d$$

Formulation

- Assume a set of regions $\{\mathcal{R}_1, \ldots \mathcal{R}_N\}$ that divide the data space
- The probability of $\mathbf{x} \in \mathcal{D}_\omega$ being in region $\mathcal{R}_j$:

$$P(\mathbf{x} \in \mathcal{R}_j) = \int_{\mathcal{R}_j} p(\mathbf{u})\, d\mathbf{u} \approx p(\mathbf{x}|\omega) \cdot V \text{ where } V \in \mathbb{R}^d$$

- Assuming regular sample distribution in data space:

$$P(\mathbf{x} \in \mathcal{R}_j) = \frac{k_j}{|\mathcal{D}|} \text{ where } k_j = |\{i : \mathbf{x}_i \in \mathcal{R}_j\}|$$

Formulation

- Assume a set of regions $\{\mathcal{R}_1, \ldots \mathcal{R}_N\}$ that divide the data space
- The probability of $\mathbf{x} \in \mathcal{D}_\omega$ being in region $\mathcal{R}_j$:

$$P(\mathbf{x} \in \mathcal{R}_j) = \int_{\mathcal{R}_j} p(\mathbf{u})\, d\mathbf{u} \approx p(\mathbf{x}|\omega) \cdot V \text{ where } V \in \mathbb{R}^d$$

- Assuming regular sample distribution in data space:

$$P(\mathbf{x} \in \mathcal{R}_j) = \frac{k_j}{|\mathcal{D}|} \text{ where } k_j = |\{i : \mathbf{x}_i \in \mathcal{R}_j\}|$$

- The density function is approximated as:

$$\hat{p}(\mathbf{x}|\omega) = \frac{k_j}{|\mathcal{D}| \cdot V}$$

Weight feature example

Weight feature example

- Unaccessible male weight distribution: $p(\mathbf{x}|\omega = \text{male}) = \mathcal{N}(80, \sqrt{40})$

Weight feature example

- Unaccessible male weight distribution: $p(\mathbf{x}|\omega = \text{male}) = \mathcal{N}(80, \sqrt{40})$
- Our real scenario:
  - $\rightarrow$ Assortment $\mathcal{D}_{\text{male}}$ with $|\mathcal{D}_{\text{male}}| = 500$ samples
  - $\rightarrow$ Approximate $p(\mathbf{x} = 91\,\mathrm{kg}\,|\,\omega = \text{male})$

Weight feature example

- Unaccessible male weight distribution: $p(\mathbf{x}|\omega = \text{male}) = \mathcal{N}(80, \sqrt{40})$
- Our real scenario:
  - $\rightarrow$ Assortment $\mathcal{D}_{\text{male}}$ with $|\mathcal{D}_{\text{male}}| = 500$ samples
  - $\rightarrow$ Approximate $p(\mathbf{x} = 91\,\mathrm{kg}\,|\,\omega = \text{male})$

Case #1

Weight feature example

- Unaccessible male weight distribution: $p(\mathbf{x}|\omega = \text{male}) = \mathcal{N}(80, \sqrt{40})$
- Our real scenario:
  - $\rightarrow$ Assortment $\mathcal{D}_{\text{male}}$ with $|\mathcal{D}_{\text{male}}| = 500$ samples
  - $\rightarrow$ Approximate $p(\mathbf{x} = 91\,\text{kg}\,|\,\omega = \text{male})$

Case #1

Volume range $[90, 95]\text{kg} \Rightarrow V = 5\text{kg}$

Weight feature example

- Unaccessible male weight distribution: $p(\mathbf{x}|\omega = \text{male}) = \mathcal{N}(80, \sqrt{40})$
- Our real scenario:
  - $\rightarrow$ Assortment $\mathcal{D}_{\text{male}}$ with $|\mathcal{D}_{\text{male}}| = 500$ samples
  - $\rightarrow$ Approximate $p(\mathbf{x} = 91\,\mathrm{kg}\,|\,\omega = \text{male})$

Case #1

Volume range $[90, 95]\mathrm{kg} \Rightarrow V = 5\mathrm{kg}$
$k_{[90,95]} = 23$

Weight feature example

- Unaccessible male weight distribution: $p(\mathbf{x}|\omega = \text{male}) = \mathcal{N}(80, \sqrt{40})$
- Our real scenario:
  - $\rightarrow$ Assortment $\mathcal{D}_{\text{male}}$ with $|\mathcal{D}_{\text{male}}| = 500$ samples
  - $\rightarrow$ Approximate $p(\mathbf{x} = 91\,\mathrm{kg} \,|\, \omega = \text{male})$

Case #1

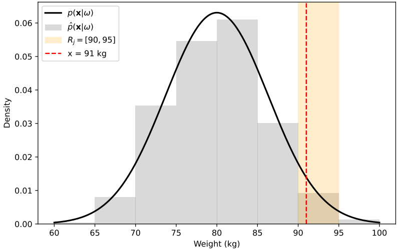
Volume range $[90, 95]\mathrm{kg} \Rightarrow V = 5\mathrm{kg}$
$k_{[90,95]} = 23$
$\hat{p}(\mathbf{x} = 91|\omega) = \frac{k_{[90,95]}}{|\mathcal{D}|\cdot V} \approx 0.009$

Weight feature example

- Unaccessible male weight distribution: $p(\mathbf{x}|\omega = \text{male}) = \mathcal{N}(80, \sqrt{40})$
- Our real scenario:
  - $\rightarrow$ Assortment $\mathcal{D}_{\text{male}}$ with $|\mathcal{D}_{\text{male}}| = 500$ samples
  - $\rightarrow$ Approximate $p(\mathbf{x} = 91\,\text{kg}\,|\,\omega = \text{male})$

Case #1

Volume range $[90, 95]\text{kg} \Rightarrow V = 5\text{kg}$
$k_{[90,95]} = 23$
$\hat{p}(\mathbf{x} = 91|\omega) = \frac{k_{[90,95]}}{|\mathcal{D}|\cdot V} \approx 0.009$



Weight feature example

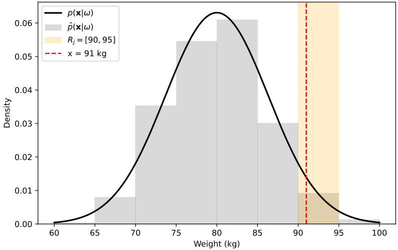
- Unaccessible male weight distribution: $p(\mathbf{x}|\omega = \text{male}) = \mathcal{N}(80, \sqrt{40})$
- Our real scenario:
  - $\rightarrow$ Assortment $\mathcal{D}_{\text{male}}$ with $|\mathcal{D}_{\text{male}}| = 500$ samples
  - $\rightarrow$ Approximate $p(\mathbf{x} = 91\,\text{kg}\,|\,\omega = \text{male})$

Case #1

Volume range $[90, 95]\text{kg} \Rightarrow V = 5\text{kg}$

$k_{[90,95]} = 23$

$\hat{p}(\mathbf{x} = 91|\omega) = \frac{k_{[90,95]}}{|\mathcal{D}|\cdot V} \approx 0.009$



Case #2

Weight feature example

- Unaccessible male weight distribution: $p(\mathbf{x}|\omega = \text{male}) = \mathcal{N}(80, \sqrt{40})$
- Our real scenario:
  - $\rightarrow$ Assortment $\mathcal{D}_{\text{male}}$ with $|\mathcal{D}_{\text{male}}| = 500$ samples
  - $\rightarrow$ Approximate $p(\mathbf{x} = 91\,\mathrm{kg}\,|\,\omega = \text{male})$

Case #1

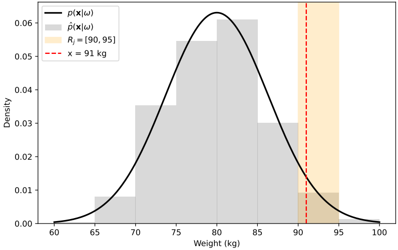
Volume range $[90, 95]\mathrm{kg} \Rightarrow V = 5\mathrm{kg}$

$k_{[90,95]} = 23$

$\hat{p}(\mathbf{x} = 91|\omega) = \frac{k_{[90,95]}}{|\mathcal{D}|\cdot V} \approx 0.009$

Case #2

Volume range $[90, 92.5]\mathrm{kg} \Rightarrow V = 2.5\mathrm{kg}$

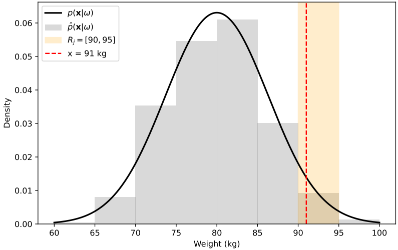


Weight feature example

- Unaccessible male weight distribution: $p(\mathbf{x}|\omega = \text{male}) = \mathcal{N}(80, \sqrt{40})$
- Our real scenario:
  - $\rightarrow$ Assortment $\mathcal{D}_{\text{male}}$ with $|\mathcal{D}_{\text{male}}| = 500$ samples
  - $\rightarrow$ Approximate $p(\mathbf{x} = 91\,\text{kg} \,|\, \omega = \text{male})$

Case #1

Volume range $[90, 95]\text{kg} \Rightarrow V = 5\text{kg}$
$k_{[90,95]} = 23$
$\hat{p}(\mathbf{x} = 91|\omega) = \frac{k_{[90,95]}}{|\mathcal{D}|\cdot V} \approx 0.009$



Case #2

Volume range $[90, 92.5]\text{kg} \Rightarrow V = 2.5\text{kg}$
$k_{[90,92.5]} = 14$

Weight feature example

- Unaccessible male weight distribution: $p(\mathbf{x}|\omega = \text{male}) = \mathcal{N}(80, \sqrt{40})$
- Our real scenario:
  - $\rightarrow$ Assortment $\mathcal{D}_{\text{male}}$ with $|\mathcal{D}_{\text{male}}| = 500$ samples
  - $\rightarrow$ Approximate $p(\mathbf{x} = 91\,\text{kg}\,|\,\omega = \text{male})$

Case #1

Volume range $[90, 95]\text{kg} \Rightarrow V = 5\text{kg}$

$k_{[90,95]} = 23$

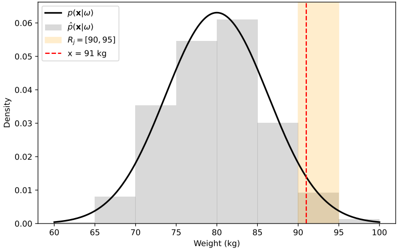
$\hat{p}(\mathbf{x} = 91|\omega) = \frac{k_{[90,95]}}{|\mathcal{D}|\cdot V} \approx 0.009$

Case #2

Volume range $[90, 92.5]\text{kg} \Rightarrow V = 2.5\text{kg}$

$k_{[90,92.5]} = 14$

$\hat{p}(\mathbf{x} = 91|\omega) = \frac{k_{[90,92.5]}}{|\mathcal{D}|\cdot V} \approx 0.011$



Weight feature example

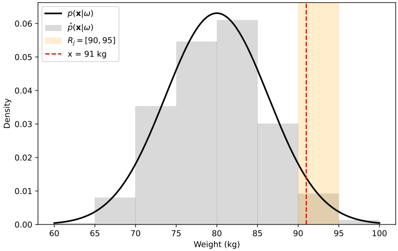
- Unaccessible male weight distribution: $p(\mathbf{x}|\omega = \text{male}) = \mathcal{N}(80, \sqrt{40})$
- Our real scenario:
  - $\rightarrow$ Assortment $\mathcal{D}_{\text{male}}$ with $|\mathcal{D}_{\text{male}}| = 500$ samples
  - $\rightarrow$ Approximate $p(\mathbf{x} = 91\,\mathrm{kg} \,|\, \omega = \text{male})$

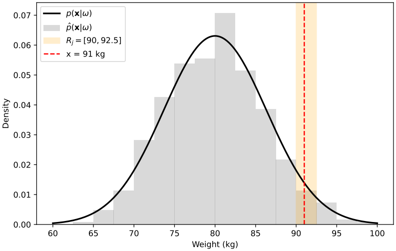
Case #1

Volume range $[90, 95]\mathrm{kg} \Rightarrow V = 5\mathrm{kg}$
$k_{[90,95]} = 23$
$\hat{p}(\mathbf{x} = 91|\omega) = \frac{k_{[90,95]}}{|\mathcal{D}|\cdot V} \approx 0.009$

Case #2

Volume range $[90, 92.5]\mathrm{kg} \Rightarrow V = 2.5\mathrm{kg}$
$k_{[90,92.5]} = 14$
$\hat{p}(\mathbf{x} = 91|\omega) = \frac{k_{[90,92.5]}}{|\mathcal{D}|\cdot V} \approx 0.011$





Weight feature example

- Unaccessible male weight distribution: $p(\mathbf{x}|\omega = \text{male}) = \mathcal{N}(80, \sqrt{40})$
- Our real scenario:
  - $\rightarrow$ Assortment $\mathcal{D}_{\text{male}}$ with $|\mathcal{D}_{\text{male}}| = 500$ samples
  - $\rightarrow$ Approximate $p(\mathbf{x} = 91\,\text{kg}\,|\,\omega = \text{male})$

Case #1

Volume range $[90, 95]\text{kg} \Rightarrow V = 5\text{kg}$

$k_{[90,95]} = 23$

$\hat{p}(\mathbf{x} = 91|\omega) = \frac{k_{[90,95]}}{|\mathcal{D}|\cdot V} \approx 0.009$

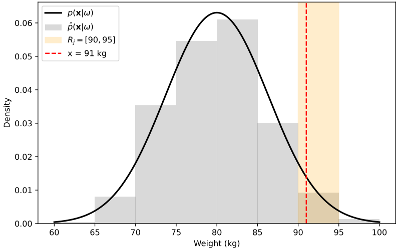
Case #2

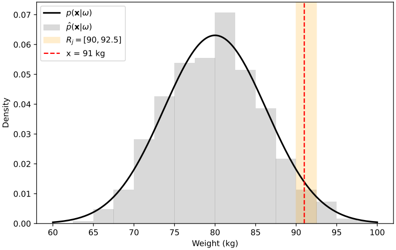
Volume range $[90, 92.5]\text{kg} \Rightarrow V = 2.5\text{kg}$

$k_{[90,92.5]} = 14$

$\hat{p}(\mathbf{x} = 91|\omega) = \frac{k_{[90,92.5]}}{|\mathcal{D}|\cdot V} \approx 0.011$

$p(\mathbf{x} = 91\,\text{kg}\,|\,\omega = \text{male}) = 0.014$





Classification example

- Classification task: *male* VS *female*:

Classification example

- Classification task: *male* VS *female*:
  - $\rightarrow$ $p(\mathbf{x}|\omega = \text{male}) = \mathcal{N}(80, \sqrt{40})$
  - $\rightarrow$ $p(\mathbf{x}|\omega = \text{female}) = \mathcal{N}(60, \sqrt{20})$

Classification example

- Classification task: *male* VS *female*:
  - $\rightarrow$ $p(\mathbf{x}|\omega = \text{male}) = \mathcal{N}(80, \sqrt{40})$
  - $\rightarrow$ $p(\mathbf{x}|\omega = \text{female}) = \mathcal{N}(60, \sqrt{20})$
- What we have:
  - $\rightarrow$ $\mathcal{D}_{\text{male}}$ assortment with $|\mathcal{D}_{\text{male}}| = 450$ samples
  - $\rightarrow$ $\mathcal{D}_{\text{female}}$ assortment with $|\mathcal{D}_{\text{female}}| = 550$ samples

Classification example

- Classification task: *male* VS *female*:
  - $\rightarrow$ $p(\mathbf{x}|\omega = \text{male}) = \mathcal{N}(80, \sqrt{40})$
  - $\rightarrow$ $p(\mathbf{x}|\omega = \text{female}) = \mathcal{N}(60, \sqrt{20})$
- What we have:
  - $\rightarrow$ $\mathcal{D}_{\text{male}}$ assortment with $|\mathcal{D}_{\text{male}}| = 450$ samples
  - $\rightarrow$ $\mathcal{D}_{\text{female}}$ assortment with $|\mathcal{D}_{\text{female}}| = 550$ samples
- What we need: Most likely gender when $\mathbf{x} = 72\text{kg}$

Classification example

- Classification task: *male* VS *female*:
  - $\rightarrow$ $p(\mathbf{x}|\omega = \text{male}) = \mathcal{N}(80, \sqrt{40})$
  - $\rightarrow$ $p(\mathbf{x}|\omega = \text{female}) = \mathcal{N}(60, \sqrt{20})$
- What we have:
  - $\rightarrow$ $\mathcal{D}_{\text{male}}$ assortment with $|\mathcal{D}_{\text{male}}| = 450$ samples
  - $\rightarrow$ $\mathcal{D}_{\text{female}}$ assortment with $|\mathcal{D}_{\text{female}}| = 550$ samples
- What we need: Most likely gender when $\mathbf{x} = 72\text{kg}$

<u>Male</u>

Volume $V = 5\text{kg}$

$k_{[70,75]} = 76$

$\hat{p}(72\,\text{kg}|\text{male}) = 0.1689$

$\text{P}(\text{male}) = {}^{450}/_{450+550} = 0.45$

Classification example

- Classification task: *male* VS *female*:
  - $\rightarrow$ $p(\mathbf{x}|\omega = \text{male}) = \mathcal{N}(80, \sqrt{40})$
  - $\rightarrow$ $p(\mathbf{x}|\omega = \text{female}) = \mathcal{N}(60, \sqrt{20})$
- What we have:
  - $\rightarrow$ $\mathcal{D}_{\text{male}}$ assortment with $|\mathcal{D}_{\text{male}}| = 450$ samples
  - $\rightarrow$ $\mathcal{D}_{\text{female}}$ assortment with $|\mathcal{D}_{\text{female}}| = 550$ samples
- What we need: Most likely gender when $\mathbf{x} = 72\text{kg}$

Male

Volume $V = 5\text{kg}$

$k_{[70,75]} = 76$

$\hat{p}(72\,\text{kg}|\text{male}) = 0.1689$

$\text{P}(\text{male}) = {}^{450}/_{450+550} = 0.45$

Female

Volume $V = 5\text{kg}$

$k_{[70,75]} = 3$

$\hat{p}(72\,\text{kg}|\text{female}) = 0.001$

$\text{P}(\text{female}) = {}^{550}/_{450+550} = 0.55$

Classification example

- Classification task: *male* VS *female*:
  - $\rightarrow$ $p(\mathbf{x}|\omega = \text{male}) = \mathcal{N}(80, \sqrt{40})$
  - $\rightarrow$ $p(\mathbf{x}|\omega = \text{female}) = \mathcal{N}(60, \sqrt{20})$
- What we have:
  - $\rightarrow$ $\mathcal{D}_{\text{male}}$ assortment with $|\mathcal{D}_{\text{male}}| = 450$ samples
  - $\rightarrow$ $\mathcal{D}_{\text{female}}$ assortment with $|\mathcal{D}_{\text{female}}| = 550$ samples
- What we need: Most likely gender when $\mathbf{x} = 72\mathrm{kg}$

Male	**Female**
Volume $V = 5\mathrm{kg}$	Volume $V = 5\mathrm{kg}$
$k_{[70,75]} = 76$	$k_{[70,75]} = 3$
$\hat{p}(72\,\mathrm{kg}\|\text{male}) = 0.1689$	$\hat{p}(72\,\mathrm{kg}\|\text{female}) = 0.001$
$\text{P}(\text{male}) = {}^{450}/_{450+550} = 0.45$	$\text{P}(\text{female}) = {}^{550}/_{450+550} = 0.55$

$$\hat{\omega} = \underset{\omega \in \{\text{male,female}\}}{\arg\max}\ \hat{p}(\mathbf{x} = 72|\omega) \cdot P(\omega)$$

Classification example

- Classification task: *male* VS *female*:
  - $\rightarrow$ $p(\mathbf{x}|\omega = \text{male}) = \mathcal{N}(80, \sqrt{40})$
  - $\rightarrow$ $p(\mathbf{x}|\omega = \text{female}) = \mathcal{N}(60, \sqrt{20})$
- What we have:
  - $\rightarrow$ $\mathcal{D}_{\text{male}}$ assortment with $|\mathcal{D}_{\text{male}}| = 450$ samples
  - $\rightarrow$ $\mathcal{D}_{\text{female}}$ assortment with $|\mathcal{D}_{\text{female}}| = 550$ samples
- What we need: Most likely gender when $\mathbf{x} = 72\text{kg}$

Male

Volume $V = 5\text{kg}$

$k_{[70,75]} = 76$

$\hat{p}(72\,\text{kg}|\text{male}) = 0.1689$

$\text{P}(\text{male}) = {}^{450}/_{450+550} = 0.45$

Female

Volume $V = 5\text{kg}$

$k_{[70,75]} = 3$

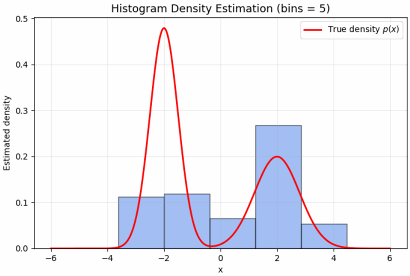
$\hat{p}(72\,\text{kg}|\text{female}) = 0.001$

$\text{P}(\text{female}) = {}^{550}/_{450+550} = 0.55$

$$\hat{\omega} = \underset{\omega \in \{\text{male},\text{female}\}}{\arg\max} \; \hat{p}(\mathbf{x} = 72|\omega) \cdot P(\omega) \Rightarrow \text{male}$$

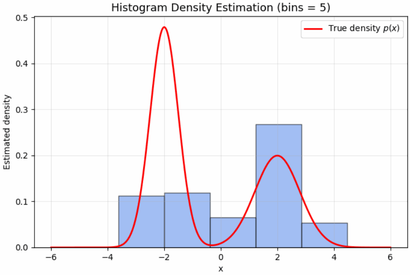
Limitations

- Does reducing the volume always improve the estimation?



Limitations

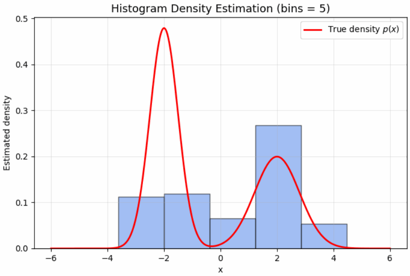
- Does reducing the volume always improve the estimation?



- Remarkably senstitive to resolution:
  - $\rightarrow$ **Large** $V \Rightarrow$ oversmoothing / high bias
  - $\rightarrow$ **Small** $V \Rightarrow$ noisy estimate / high variance

Limitations

- Does reducing the volume always improve the estimation?



- Remarkably senstitive to resolution:
  - $\rightarrow$ **Large** $V \Rightarrow$ oversmoothing / high bias
  - $\rightarrow$ **Small** $V \Rightarrow$ noisy estimate / high variance
- Also discontinuous, sensible to bin alignment, and coarse

Formulation

Parzen window estimator: Generalization of the histogram estimator

Formulation

Parzen window estimator: Generalization of the histogram estimator

- Query **x** contributes locally to the density around its own position:

$$\hat{p}(\mathbf{x}|\omega) = \frac{k_j}{|\mathcal{D}| \cdot V} \Rightarrow \frac{k_j(\mathbf{x})}{|\mathcal{D}| \cdot V}$$

Formulation

Parzen window estimator: Generalization of the histogram estimator

- Query **x** contributes locally to the density around its own position:

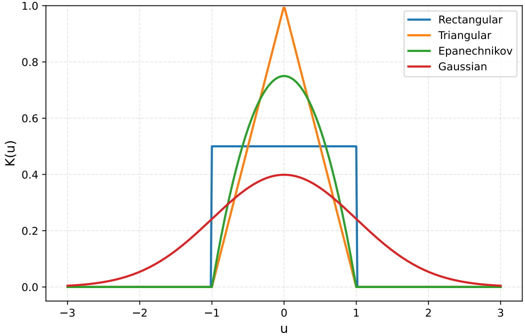
$$\hat{p}(\mathbf{x}|\omega) = \frac{k_j}{|\mathcal{D}| \cdot V} \Rightarrow \frac{k_j(\mathbf{x})}{|\mathcal{D}| \cdot V}$$

- Instead of rigid bins, **kernel** function around **x**:

$$\hat{p}(\mathbf{x}|\omega) = \frac{1}{|\mathcal{D}| \cdot h^d} \sum_{i=1}^{|\mathcal{D}|} K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right)$$

$h \rightarrow$ window width

Window shapes



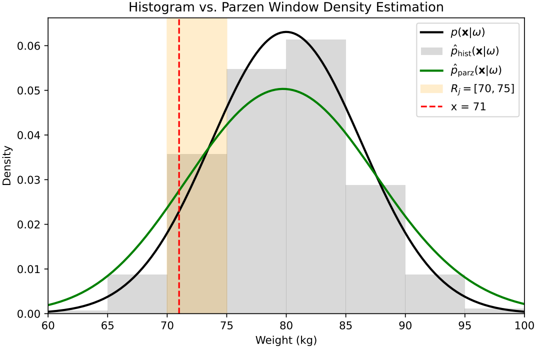
$$K_{\text{rec}}(u) = \begin{cases} 1/2 & |u| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

$$K_{\text{epa}}(u) = \begin{cases} \frac{3}{4}(1 - u^2) & |u| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

$$K_{\text{tri}}(u) = \begin{cases} 1 - |u| & |u| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

$$K_{\text{gau}}(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$$

Formulation



Issues

- **Main limitation**: fixed window width h regardless of $\mathbf{x}$

Issues

- **Main limitation**: fixed window width h regardless of $\mathbf{x}$

- Same neighborhood size regardless of (local) data density/sparsity:
  - $\rightarrow$ High density $\Rightarrow$ too many samples $\Rightarrow$ over-smoothed estimation
  - $\rightarrow$ Low density $\Rightarrow$ too few samples $\Rightarrow$ under-smoothed estimation

Formulation

k_n-**Nearest Neighbor estimator**: Addresses the limitations in Parzen

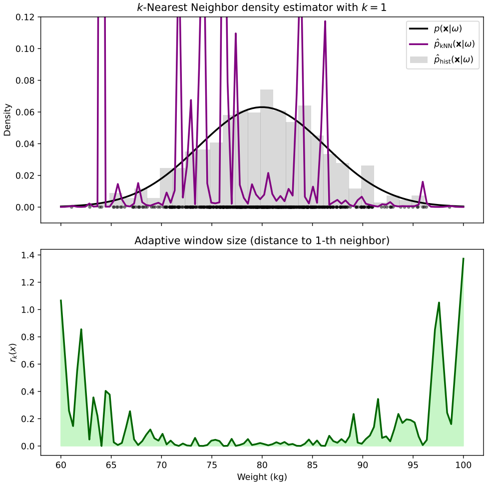
Formulation

k_n-**Nearest Neighbor estimator**: Addresses the limitations in Parzen

- ✗ Does not fix window width h
- ✓ Fixes number of samples k and lets the window volume vary

$$\hat{p}(\mathbf{x}|\omega) = \frac{k_j}{|\mathcal{D}| \cdot V} \Rightarrow \frac{k}{|\mathcal{D}| \cdot V(\mathbf{x})}$$

Formulation



Posterior probability

Posterior probability

- Assortment $\mathcal{D}$ from an unknown distribution
  - $\rightarrow$ Subset $\mathcal{D}_\omega = \{(\mathbf{x}_i, y_i) \in \mathcal{D} : y_i = \omega\}_{i=1}^{|\mathcal{D}|}$ with class ω

Posterior probability

- Assortment $\mathcal{D}$ from an unknown distribution
  - $\rightarrow$ Subset $\mathcal{D}_\omega = \{(\mathbf{x}_i, y_i) \in \mathcal{D} : y_i = \omega\}_{i=1}^{|\mathcal{D}|}$ with class ω
- Estimation using k neighbors
  - $\rightarrow$ k_ω: number of neighbors with class ω

Posterior probability

- Assortment $\mathcal{D}$ from an unknown distribution
  - $\rightarrow$ Subset $\mathcal{D}_\omega = \{(\mathbf{x}_i, y_i) \in \mathcal{D} : y_i = \omega\}_{i=1}^{|\mathcal{D}|}$ with class ω
- Estimation using k neighbors
  - $\rightarrow$ k_ω: number of neighbors with class ω

$$\hat{P}(\omega|\mathbf{x}) = \frac{\hat{p}(\mathbf{x}|\omega) \cdot P(\omega)}{\hat{p}(\mathbf{x})}$$

Posterior probability

- Assortment $\mathcal{D}$ from an unknown distribution
  - $\rightarrow$ Subset $\mathcal{D}_\omega = \{(\mathbf{x}_i, y_i) \in \mathcal{D} : y_i = \omega\}_{i=1}^{|\mathcal{D}|}$ with class ω
- Estimation using k neighbors
  - $\rightarrow$ k_ω: number of neighbors with class ω

$$\hat{P}(\omega|\mathbf{x}) = \frac{\hat{p}(\mathbf{x}|\omega) \cdot P(\omega)}{\hat{p}(\mathbf{x})} \approx \frac{\frac{k_\omega/|\mathcal{D}_\omega|}{V(\mathbf{x})} \cdot \frac{|\mathcal{D}_\omega|}{|\mathcal{D}|}}{\frac{k/|\mathcal{D}|}{V(\mathbf{x})}}$$

Posterior probability

- Assortment $\mathcal{D}$ from an unknown distribution
  - $\rightarrow$ Subset $\mathcal{D}_\omega = \{(\mathbf{x}_i, y_i) \in \mathcal{D} : y_i = \omega\}_{i=1}^{|\mathcal{D}|}$ with class ω
- Estimation using k neighbors
  - $\rightarrow$ k_ω: number of neighbors with class ω

$$\hat{P}(\omega|\mathbf{x}) = \frac{\hat{p}(\mathbf{x}|\omega) \cdot P(\omega)}{\hat{p}(\mathbf{x})} \approx \frac{\frac{k_\omega/|\mathcal{D}_\omega|}{V(\mathbf{x})} \cdot \frac{|\mathcal{D}_\omega|}{|\mathcal{D}|}}{\frac{k/|\mathcal{D}|}{V(\mathbf{x})}} = \frac{k_\omega}{k}$$

Issues

- Hyperparameter k remarkably influences the estimation:
  - $\rightarrow$ $k \downarrow$: noisy estimate
  - $\rightarrow$ $k \uparrow$: oversmoothed estimate
- Computational inefficiency
- Lack of smoothness / discontinuity
  - $\rightarrow$ Piecewise estimate

Comparative summary of the density estimators

Estimator	**Volume size**	**Number of points**	**Advantages**	**Disadvantages**
Histogram	Fixed (bins)	Variable	Simple	Discontinuous
Parzen	Fixed (width h)	Variable	Smooth	Not adaptive
k_n-NN	Variable	Fixed	Adaptive	Not smooth, costly