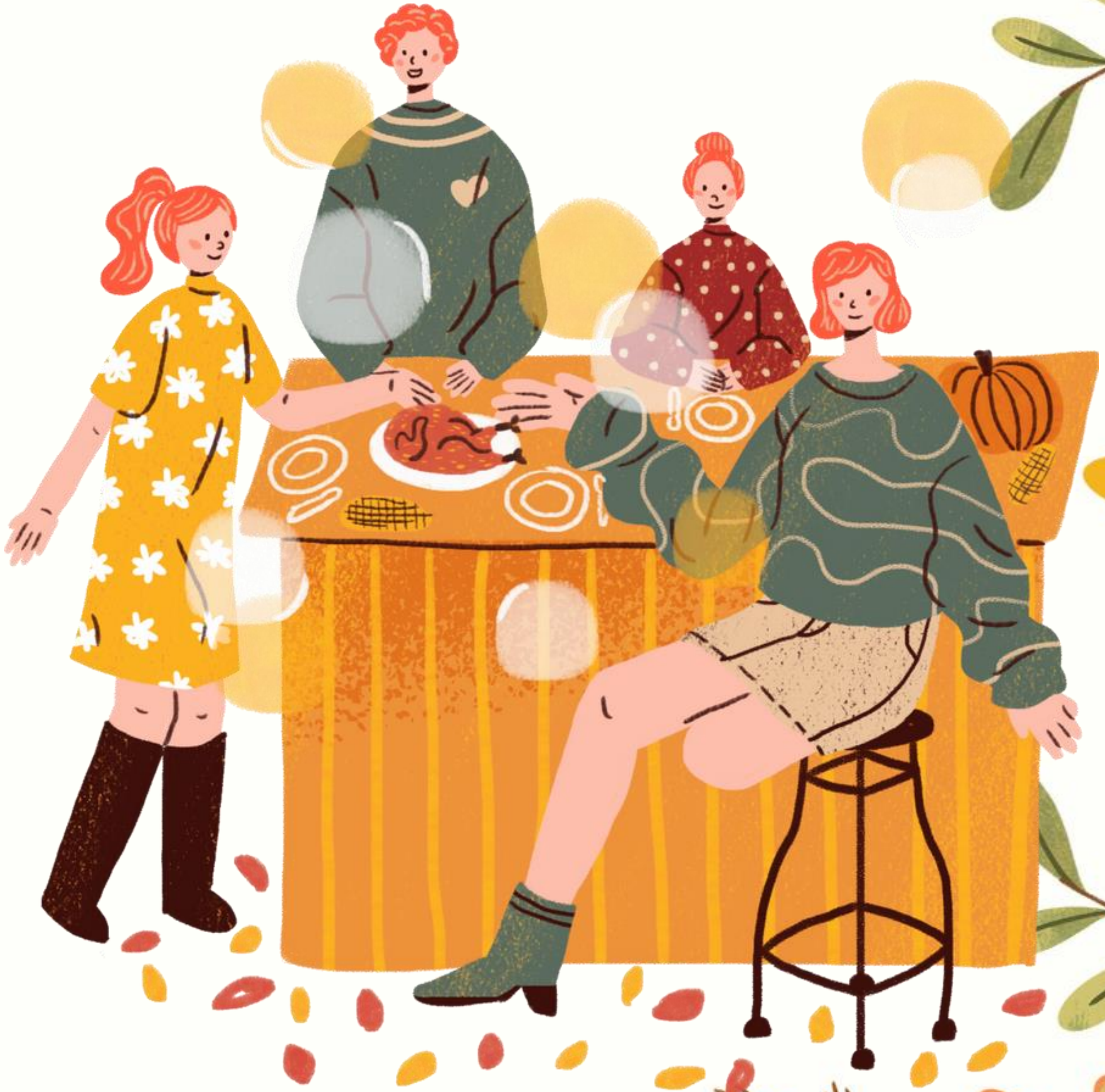
Introduction to Gaussian Processes

Yijing Chen

OVERVIEW

- Fundamental Mathematical Concepts
- Gaussian Process
- Reproducing Kernel Hilbert Space (RKHS)
- Kernel Ridge Regression
- Gaussian Process Regression
- Connection between GP Regression and KRR



Multivariate Gaussian Distribution

- **Block Form**

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{matrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{matrix}\right)$$

- **Marginally**

$$X_1 \sim \mathrm{N}(\mu_1, \Sigma_{11}), \qquad X_2 \sim \mathrm{N}(\mu_2, \Sigma_{22})$$

- **Conditionally**

$$X_1 \mid X_2 = \mathrm{x}_2 \sim \mathrm{N}(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

Gaussian Process

- **Definition**

$$f \sim GP(\mathrm{m}, \mathrm{k})$$

- **Mean Function**

$$m: \mathbb{R} \to \mathbb{R}, \quad m(x) = \mathbb{E}[f(\mathrm{x})]$$

- **Covariance (kernel) Function**

$$k: \mathbb{R} \times \mathbb{R} \to \mathbb{R}, \qquad k(x, x') = Cov(f(\mathrm{x}), f(x'))$$

Gaussian Process

For any finite set of inputs

$$X = (x_1, \ldots, x_n)$$

We have

$$f_X = (f(x_1), \ldots, f(x_n))^T \sim N(m_X, k_{XX})$$

with

$$m_X = \begin{pmatrix} m(x_1) \\ \vdots \\ m(x_n) \end{pmatrix} \qquad k_{XX} = \begin{pmatrix} k(x_1, x_1) & \cdots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \cdots & k(x_n, x_n) \end{pmatrix}$$

Reproducing Kernel Hilbert Space (RKHS)

- **Definition**

Function space $\mathcal{H}_\mathrm{k}$ associated with a kernel k

- **Reproducing property**

$$f(x) = < f, k(x,\cdot) >_{\mathcal{H}_\mathrm{k}}$$

Kernel Ridge Regression

Setting

- **Data:** $(x_i, y_i)_{i=1}^n, \qquad x_i \in \mathbb{R}^d,\ y_i \in \mathbb{R}$

- **Optimization problem** $$\hat{f} = \arg\min_{f \in \mathcal{H}_k} \frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \|f\|_{\mathcal{H}_k}^2$$

- **Closed-form solution** $$\hat{f}(x) = k_{xX}(k_{XX} + n\lambda I_n)^{-1} Y = \sum_{i=1}^n \alpha_i k(x, x_i), \quad x \in \mathcal{X},$$

where

$$(\alpha_1, \cdots, \alpha_n)^T = (k_{XX} + n\lambda I_n)^{-1} Y \in \mathbb{R}^n.$$

Gaussian Process Regression

- **Model**

Assume :

$$f \sim GP(\mathrm{m}, \mathrm{k}), \quad y_i = f(x_i) + \varepsilon_i, \quad \varepsilon_i \sim \mathrm{N}(0, \sigma^2)$$

- **Posterior Distribution** :

Given training inputs: $\mathrm{X} = (x_1, \dots, x_n)^T, \quad \mathrm{Y} = (y_1, \dots, y_n)^T$

posterior is again a Gaussian Process:

$$f \,|\, Y \sim \mathrm{GP}\,(\bar{m}, \bar{k})$$

- **Posterior Mean** :

$$\bar{m}(x) = m(x) + k_{xX}(k_{XX} + \sigma^2 I_n)^{-1}(Y - m_X)$$

- **Posterior Covariance** :

$$\bar{k}(x, x') = k(x, x') - k_{xX}(k_{XX} + \sigma^2 I_n)^{-1} k_{Xx'},$$

where

$$k_{Xx} = k_{xX}^T = (k(x_1, x), \cdots, k(x_n, x))^T.$$

Connection between GP Regression and KRR

GP posterior mean = Kernel Ridge Regression

- KRR Prediction

$$\hat{f}(x_*) = k_{x_* X}\left(k_{XX} + \lambda I_n\right)^{-1} Y.$$

- GP posterior mean

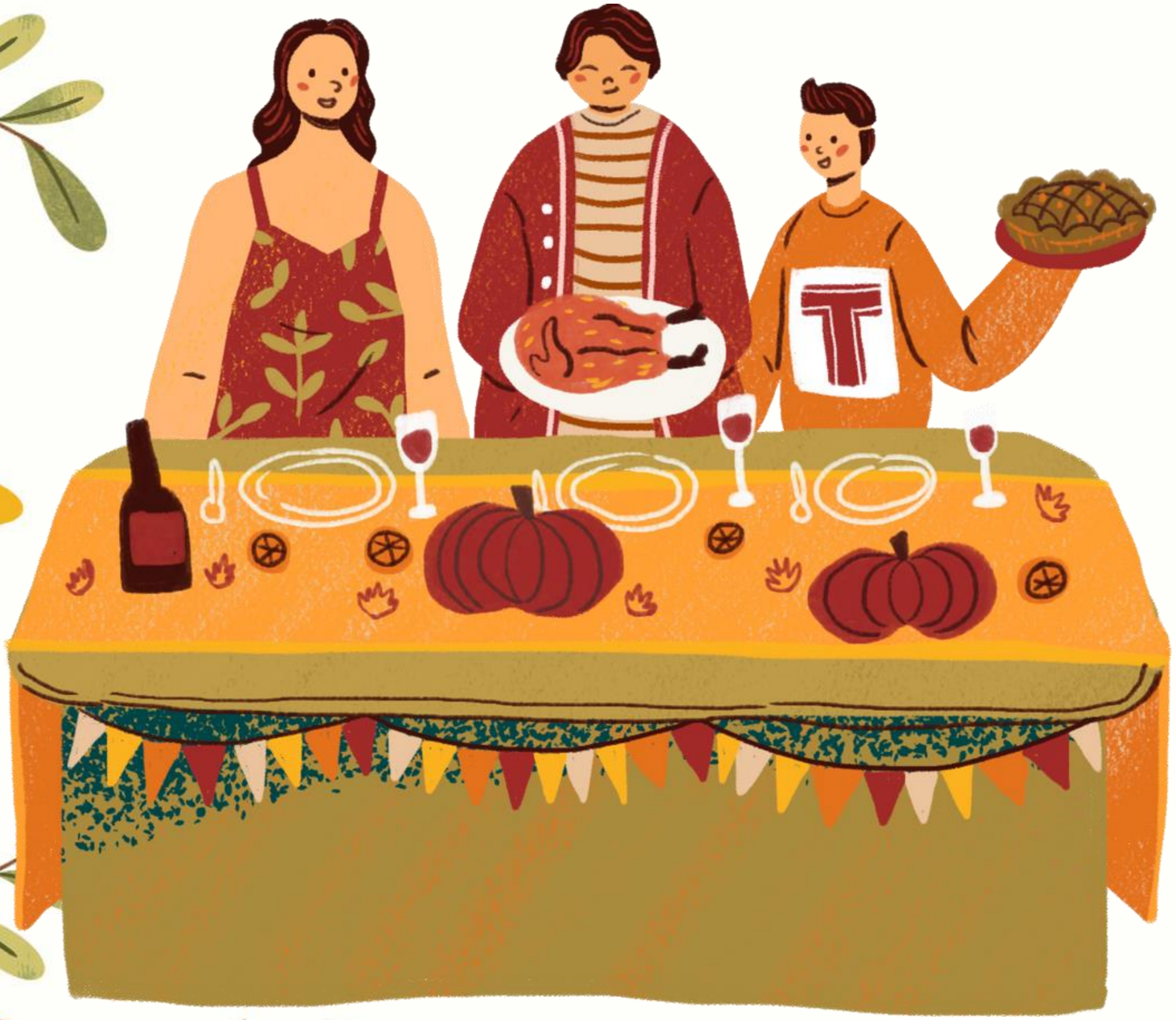
$$\bar{m}(x_*) = m(x_*) + k_{x_* X}\left(k_{XX} + \sigma^2 I_n\right)^{-1}(Y - m_X).$$

- Equivalence(zero-mean GP, $\lambda = \sigma^2$/n)

$$n\lambda = \sigma^2$$

$$\hat{f}(x_*) = \bar{m}(x_*).$$

Question?



THANK YOU!