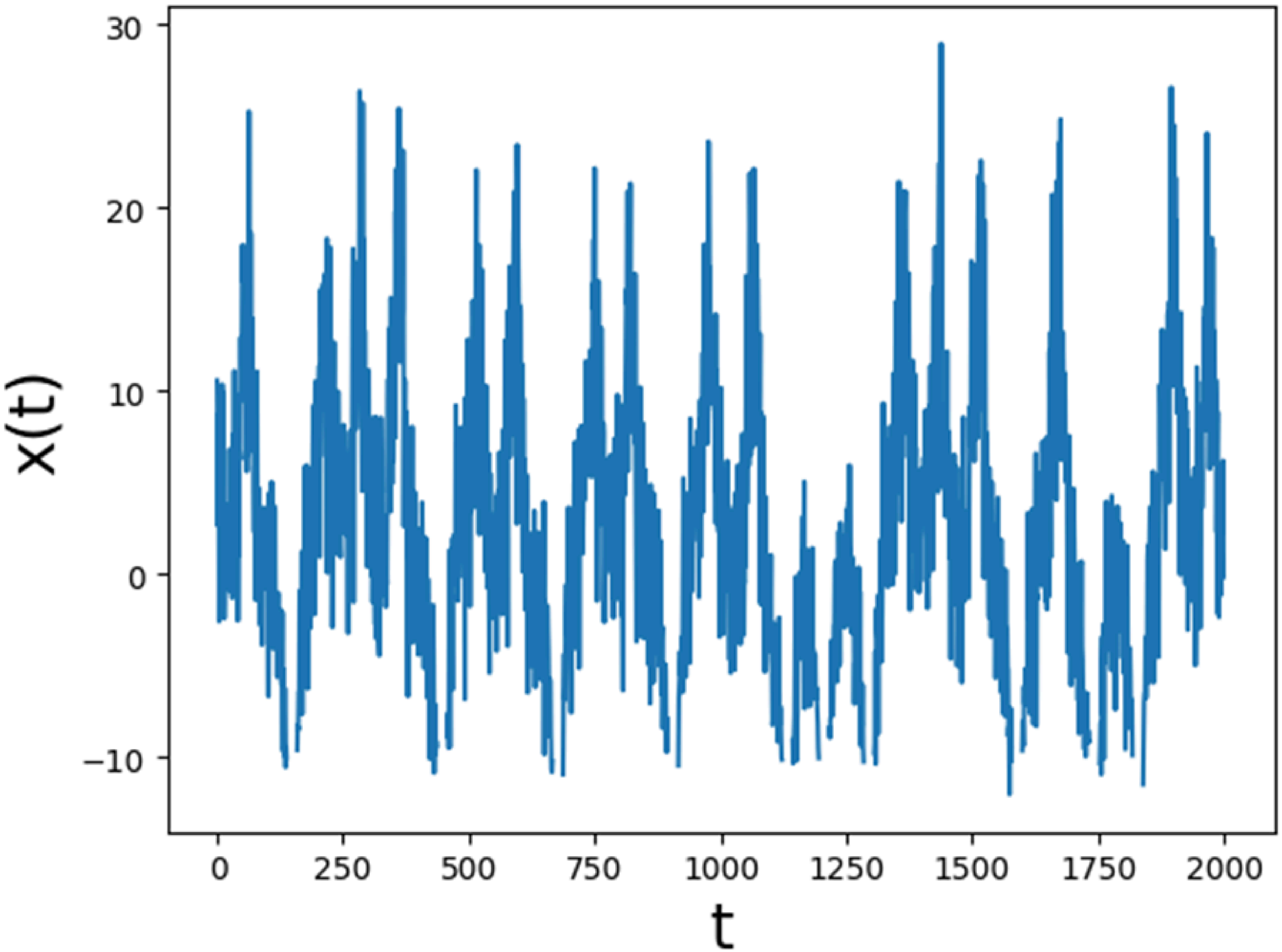
THE STUDY OF UNCERTAINTY IN FORECASTING TIME SERIES

UE D: DEEP LEARNING

Benjamin Richards
Cauê Caviglioni

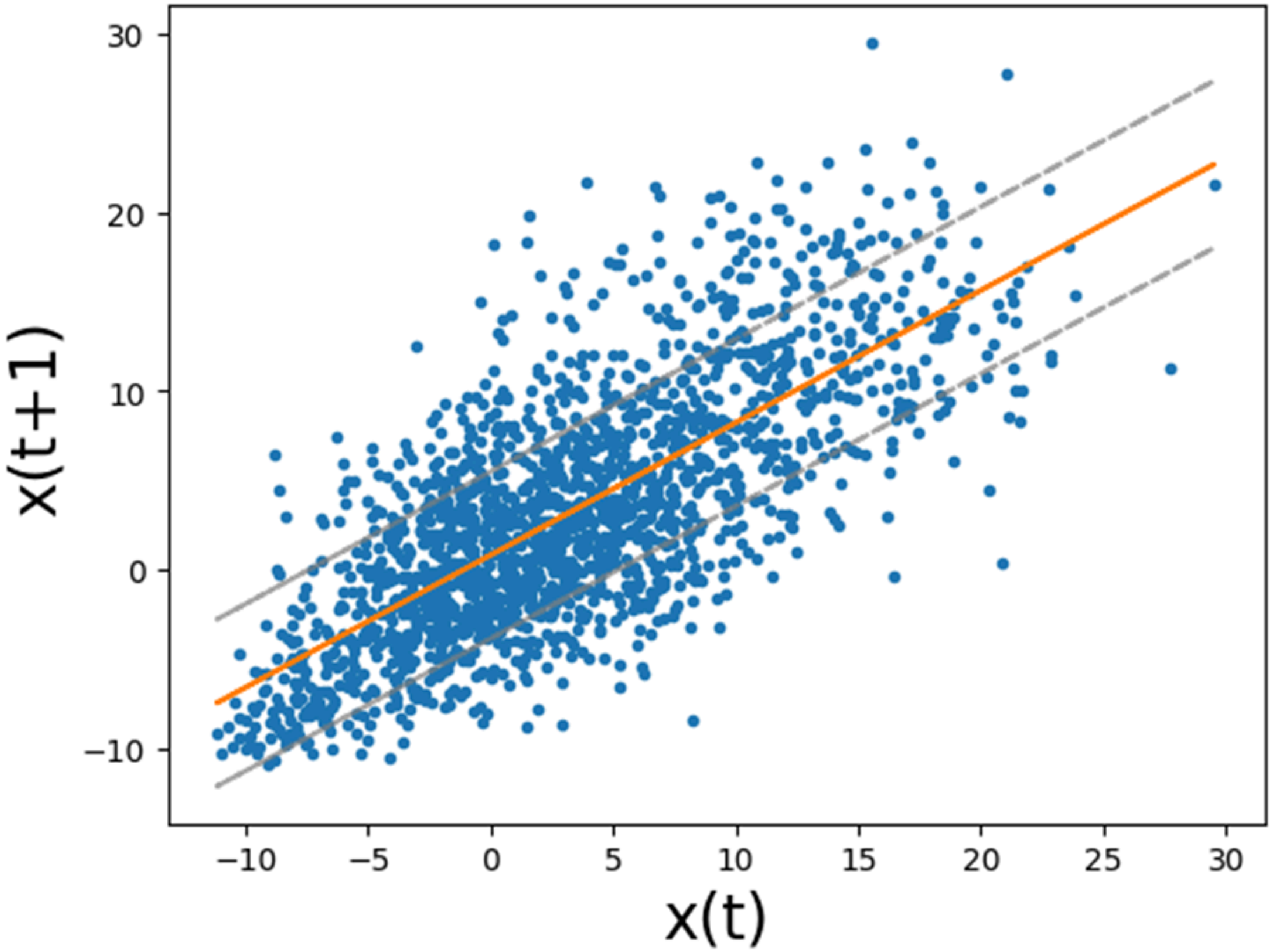
The problem



Auto Regressive Models AR(1)

$$x_t = \theta_1 x_{t-1} + \theta_0 + \epsilon_t$$

Auto Regressive Models



$$x_t = \theta_1 x_{t-1} + \theta_0 + \epsilon_t$$

Approaches

1) The parametric estimation of predicted values along with a constant standard deviation.

2) The parametric estimation of predicted values along with a non-constant standard deviation.

3) The non-parametric estimation of predicted values along with a non-constant standard deviation.

Approach 1

- Probability of observations

$$P(x_t \mid \hat{x}_t, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_t - \hat{x}_t)^2}{2\sigma^2}\right)$$

- Cost Function

$$\begin{aligned}
\mathcal{L} &= \prod_{i=1}^{n} f(x_i \mid \hat{x}_i, \sigma^2) \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \hat{x}_i)^2}{2\sigma^2}\right) \\
&= -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2} \sum_{i=1}^{n} \frac{(x_i - \hat{x_i})^2}{2\sigma^2}
\end{aligned}$$

Approach 1

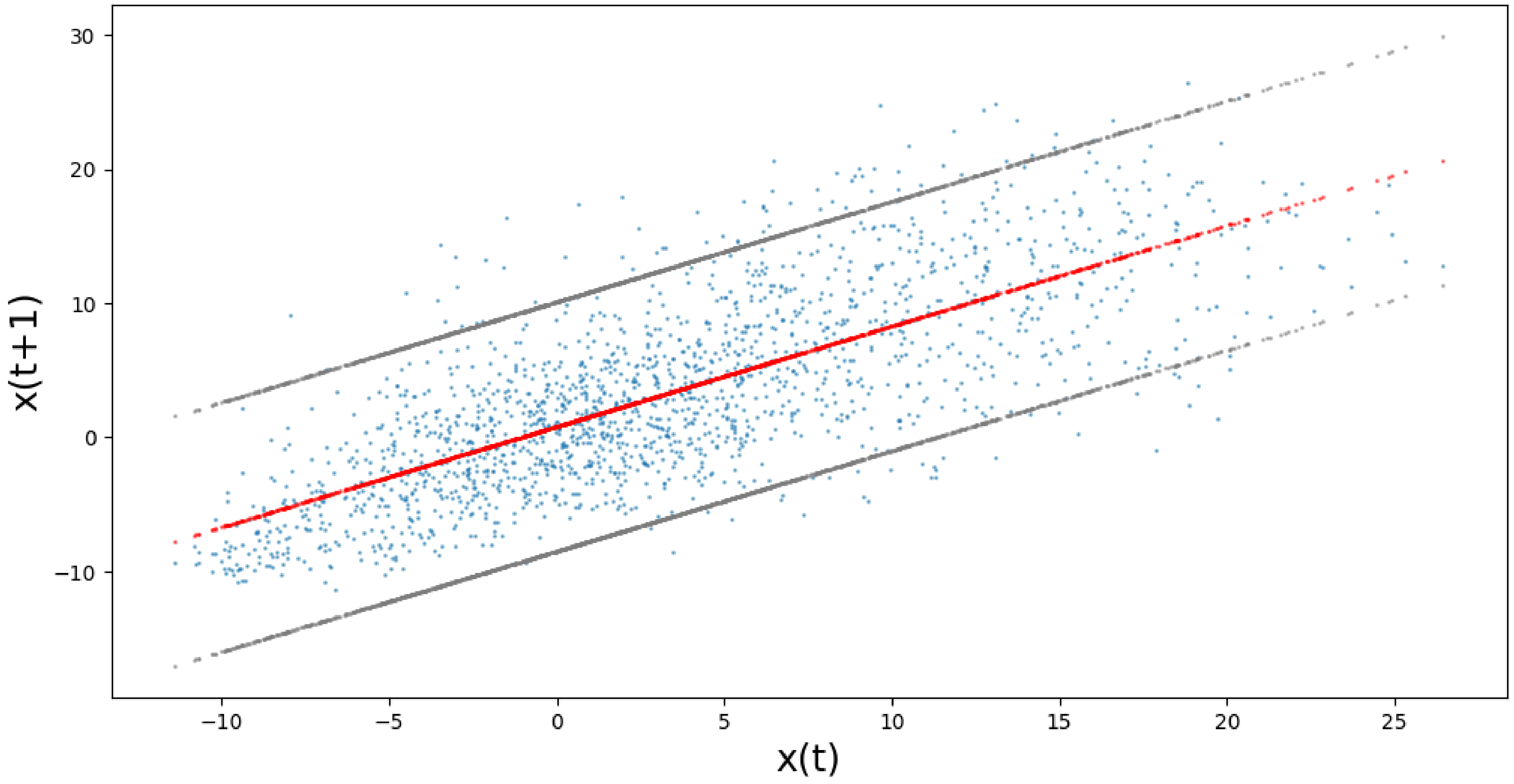
- Architecture

- Linear(1,1) for the coefficients of regression line

- Variance as a training parameter

- The models are trained together

- Results



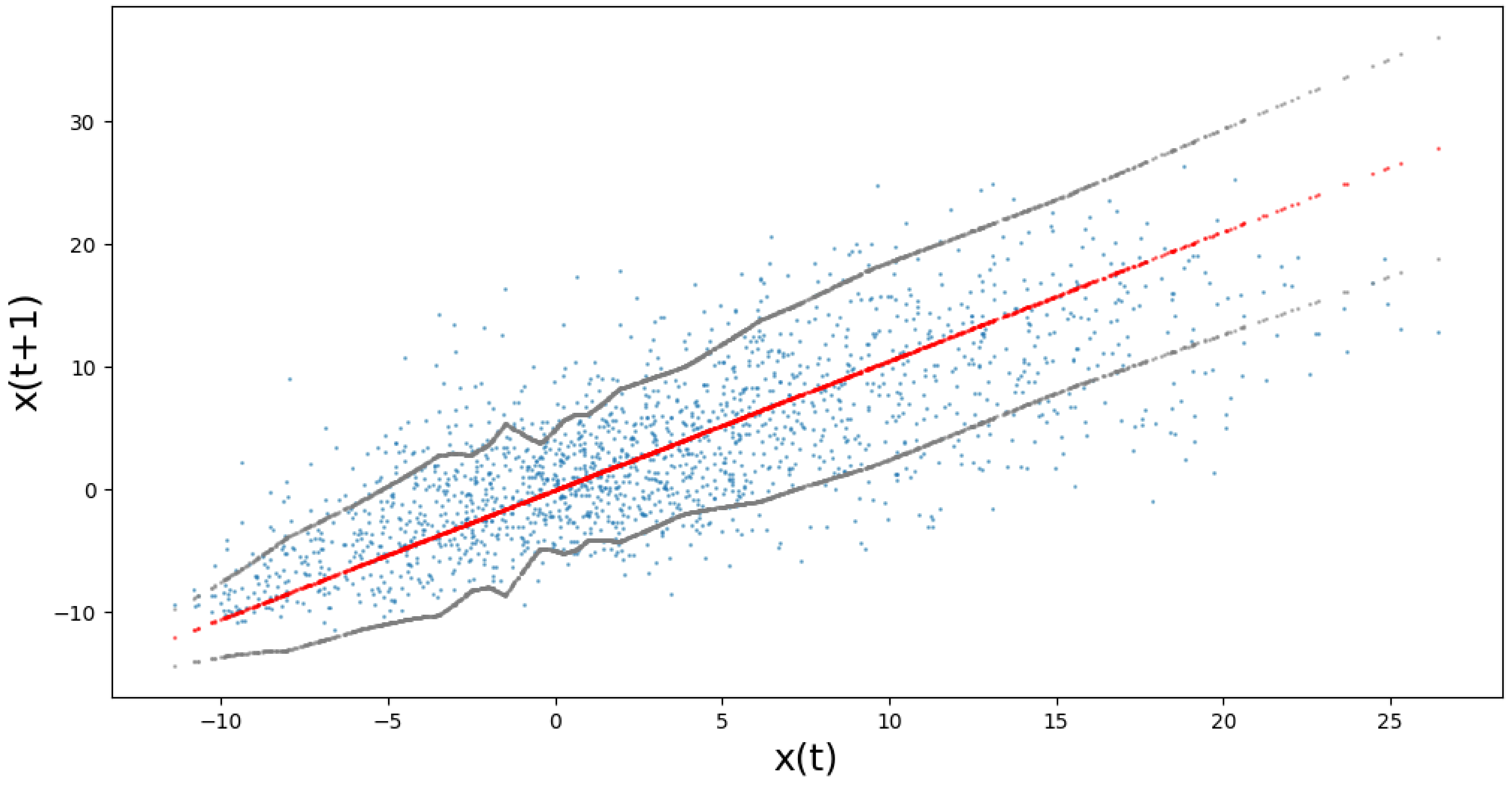
Approach 2

- Assuming that the standard deviation depends on time
- Need to slightly modify lost function

$$
\begin{aligned}
\mathcal{L} &= \prod_{i=1}^{n} f(x_i \mid \hat{x}_i, \sigma^2(i)) \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma(i)^2}} \exp\left(-\frac{(x_i - \hat{x}_i)^2}{2\sigma(i)^2}\right) \\
&= -\sum_{i=1}^{n} \frac{1}{2} \ln(2\pi\sigma(i)^2) - \frac{1}{2} \sum_{i=1}^{n} \frac{(x_i - \hat{x}_i)^2}{2\sigma(i)^2}
\end{aligned}
$$

Approach 2

- Linear(1,1) coefficients of regression line
- Two layers for variance:
    - Linear(1, 256)
    - Linear(256, 1)
    - ReLU activation

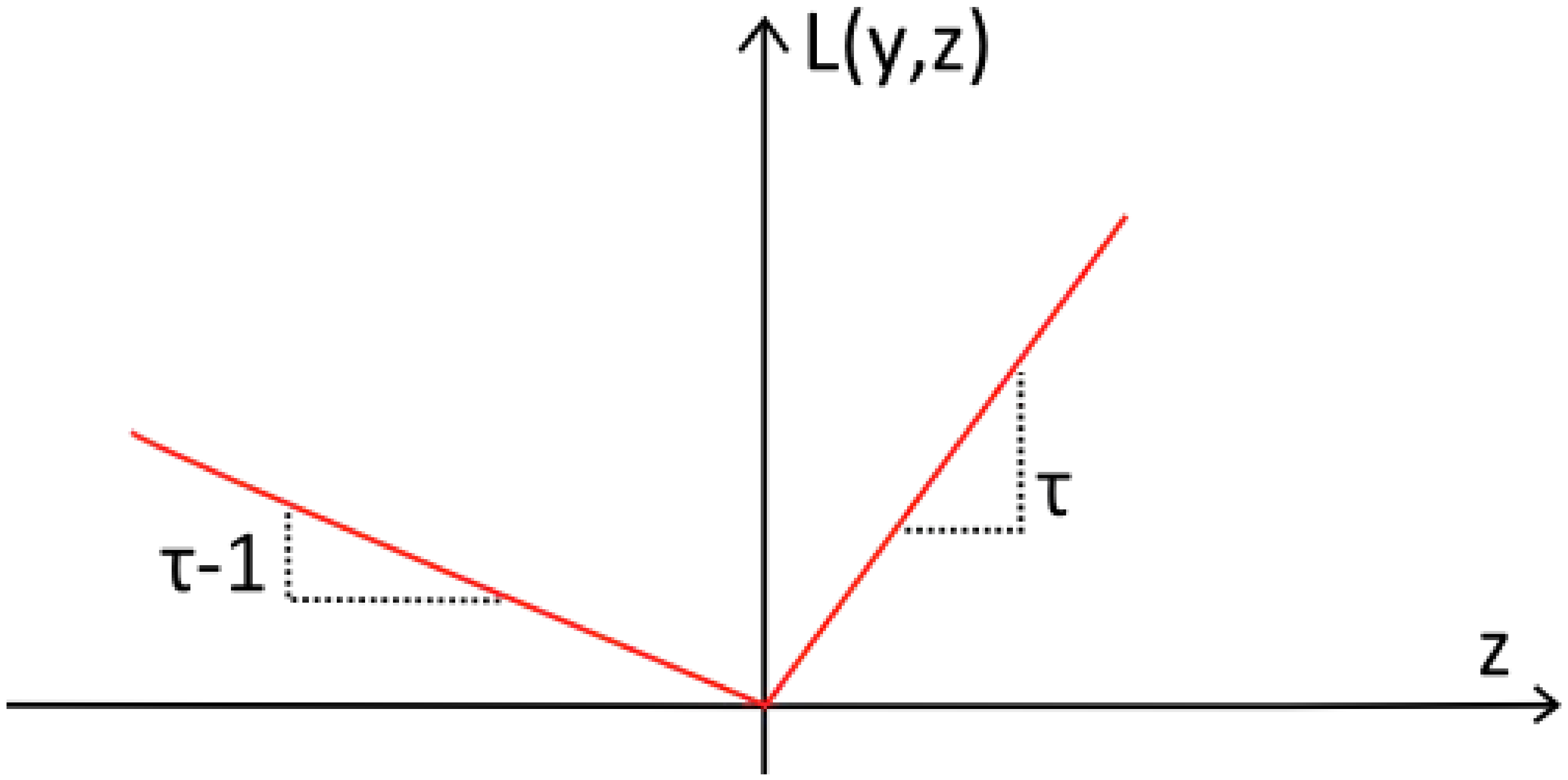


Approach 3

- Non-parametric approach requires completely different loss function
- Estimation of quantiles
- Pinball loss

$$
\begin{aligned}
L_\tau(y, z) &= (y - z)\tau & si \;\; y \geq z \\
&= (z - y)(1 - \tau) & si \;\; z > y
\end{aligned}
$$

Approach 3



$$\text{Perte totale}(z) = \sum_{i=1}^{n} L_\tau(y_i, z)$$
$$= \tau \sum_{i:y_i \geq z} (y_i - z) + (1-\tau) \sum_{i:y_i < z} (z - y_i)$$

Approach 3

- 5%, 50% and 95% quantiles estimated
- Each quantile has two layers :
  - Linear(1, 64)
  - Linear(64, 1)
  - Softmax activation