8.1 Sequence models

Lecture based on "Dive into Deep Learning" **http://D2L.AI** (Zhang et al., 2020)

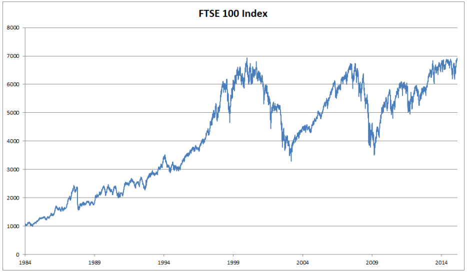
Prof. Dr. Christoph Lippert

Digital Health & Machine Learning

Example (Stock prices)

Let's denote the prices by $x_t \geq 0$, i.e. at time $t \in \mathbb{N}$ we observe price x_t.
Predict the stock market price x_t on day t via

$$x_t \sim p(x_t | x_{t-1}, \dots x_1).$$



Problem:
How do we train a model to predict x_t?

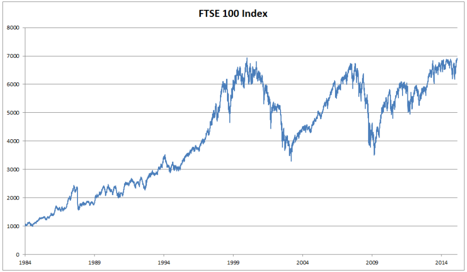
- Use historical observations to predict the next observation given the ones up to right now.
- Values of x_t change, but the dynamics of the time series are typically assumed to not change (**stationary**).
- Estimate of the entire time series via

$$p(x_1, \dots x_T) = \prod_{t=1}^{T} p(x_t | x_{t-1}, \dots x_1).$$

Example (Stock prices)

Let's denote the prices by $x_t \geq 0$, i.e. at time $t \in \mathbb{N}$ we observe price x_t.
Predict the stock market price x_t on day t via

$$x_t \sim p(x_t | x_{t-1}, \dots x_1).$$



Problem:
The number of inputs, $x_{t-1}, \dots x_1$ varies with t.

Solution 1: **Autoregressive** models

- Assume that the complete sequence $x_{t-1}, \dots x_1$ isn't necessary.
- Assume timespan τ sufficient and only use $x_{t-1}, \dots x_{t-\tau}$ observations.
- Number of arguments is always the same, for $t > \tau$.

Markov model

- In an autoregressive model we use only $(x_{t-1}, \dots x_{t-\tau})$ instead of $(x_{t-1}, \dots x_1)$ to estimate x_t.
- If $\tau = 1$, we have a **first order** Markov model and $p(x)$ is given by

$$p(x_1, \dots x_T) = \prod_{t=1}^{T} p(x_t | x_{t-1}).$$

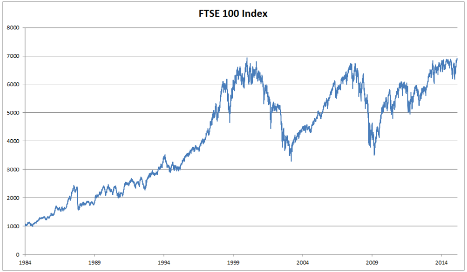
- If x_t is discrete, **dynamic programming** can be used to compute values along the chain exactly.
  - $x_{t+1} | x_{t-1}$ is computed from

$$p(x_{t+1} | x_{t-1}) = \sum_{x_t} p(x_{t+1} | x_t) p(x_t | x_{t-1})$$

Example (Stock prices)

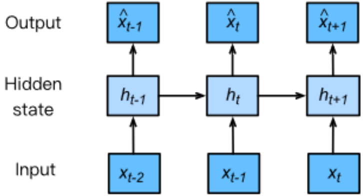
Let's denote the prices by $x_t \geq 0$, i.e. at time $t \in \mathbb{N}$ we observe price x_t.
Predict the stock market price x_t on day t via

$$x_t \sim p(x_t | x_{t-1}, \ldots x_1).$$



Problem:
The number of inputs, $x_{t-1}, \ldots x_1$ varies with on t.

Solution 2: **Latent Autoregressive** models



- introduce summary variable h_t of the past observations
- estimate $x_t | x_{t-1}, h_{t-1}$

Note on the direction of a time series

In principle, it is possible to unfold $p(x_1, \ldots x_T)$ in reverse order.

By conditioning we can always write

$$p(x_1, \ldots x_T) = \prod_{t=T}^{1} p(x_t | x_{t+1}, \ldots x_T).$$

- Usually, there exists a natural direction for the data (e.g., forward in time).
- If we change x_t, we may be able to influence x_{t+1} going forward but not the converse.
- Consequently, it ought be easier to explain $x_{t+1}|x_t$ rather than $x_t|x_{t+1}$.

  Hoyer et al., 2008 show that in some cases we can find $x_{t+1} = f(x_t) + \epsilon$ for some additive noise, whereas the converse is not true.[1]

[1] For more on **causal inference** see e.g. the book by Peters, Janzing and Schölkopf, 2015.

Summary – Sequence models

- Sequence models require specialized statistical tools for estimation.
  - **autoregressive models**
  - **latent-variable autoregressive models**
- If you have a time series, always respect the **temporal order** of the data when training, i.e. never train on future data.
- For causal models (e.g. time going forward), estimating the forward direction is typically easier than the reverse direction. i.e. we can get use simpler networks.