Deep Learning.

Today: 31/03 is overview of NN, DL.

→ 07/04 – 09/04 → Two great lectures by Ashish Tendulkar from Google.

. 14/04 → DL.

Linear model for regression / for classification → $(x_i, x_j) = K(x_i, x_j)$ → +ve semi definite

$$k(x_i, x_j) = \phi^T(x_i)\,\phi(x_j)$$

$\{x_i\}_{i \in [1000]}$ → $F(x)$ → y

$F(x)$ → $\exp(\log(\dots))$

$\min_w (y - w^T x)^2$ ⇝ poor error.

$\min_{f_\theta} (y - f_\theta(x))^2$ → θ

$\frac{\partial f_\theta}{\partial \theta}$ → GD / SGD / Adam

2014 → TF becomes public.

Here will discuss about complex relation ship between (y, x), $\because \longrightarrow y = F(x)$

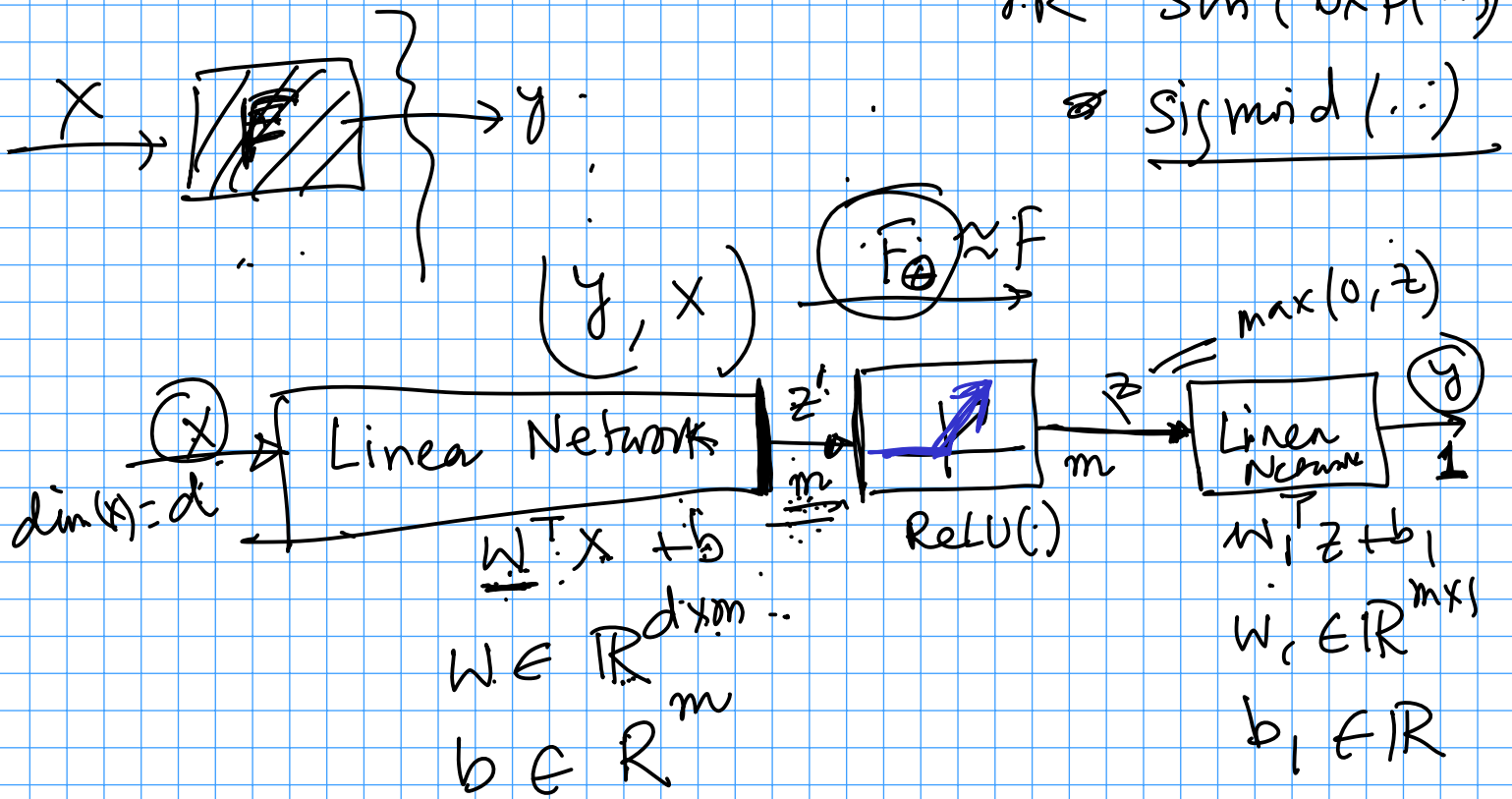
$$\min_{\theta} \left(y - F_\theta(x)\right)^2 \longrightarrow F_\theta(x) \simeq F(x) \downarrow \text{True } F.$$

Challenge

① F can be complex $\longrightarrow$ ~~TF~~; Pytorch. GD, Adam.

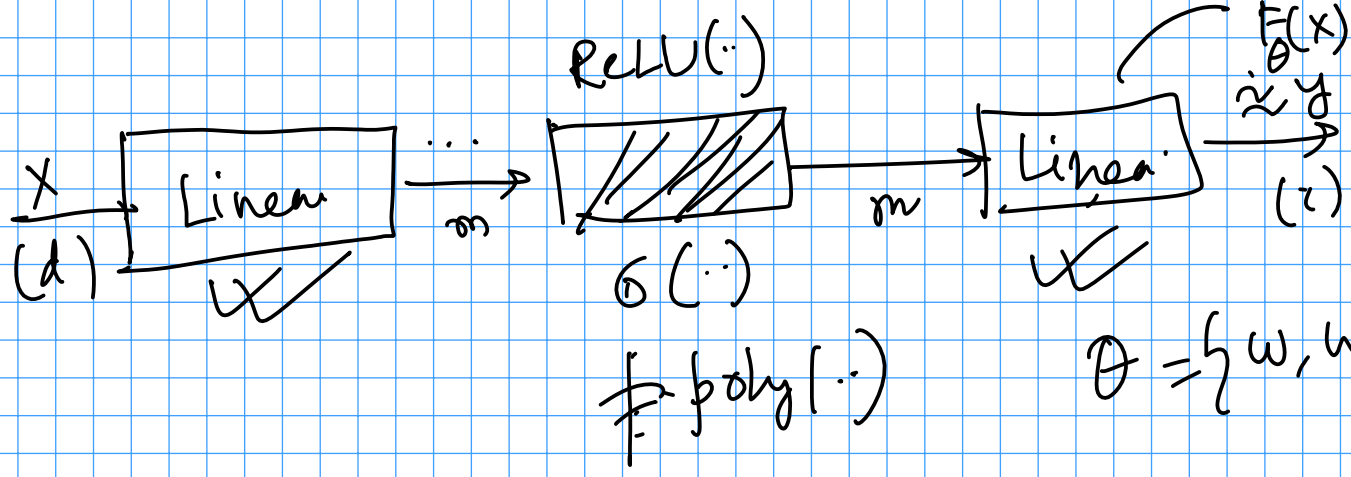
② We don't know (F).

we don't know if $F = \underline{\exp(\log(..))}$ or: $\underline{\sin(\exp(..))}$ & $\underline{\text{Sigmoid}(..)}$



$F(x) \simeq F_\theta(x)$?

Linear $\overset{m}{\downarrow}$ + ReLU $\overset{m}{\downarrow}$ + Linear $\simeq$ any nonlinearity



$$F_\theta(X) \simeq F(X)$$

$(X, y) \longrightarrow$ Linear ReLU Linear $(y \in \mathbb{R})$ if y is a continuous number.
Linear σ Linear

(X, y): X → feature, y → labels

$\longrightarrow$ Linear ReLU Lin. $\to \sigma(\cdot)$ $\quad y \in \{+1, -1\}$
Linear σ Linear $\to \sigma(\cdot)$

$X_1, X_2, X_3 \ldots X_n$ sequence

$y_1, y_2, y_3 \ldots y_n$

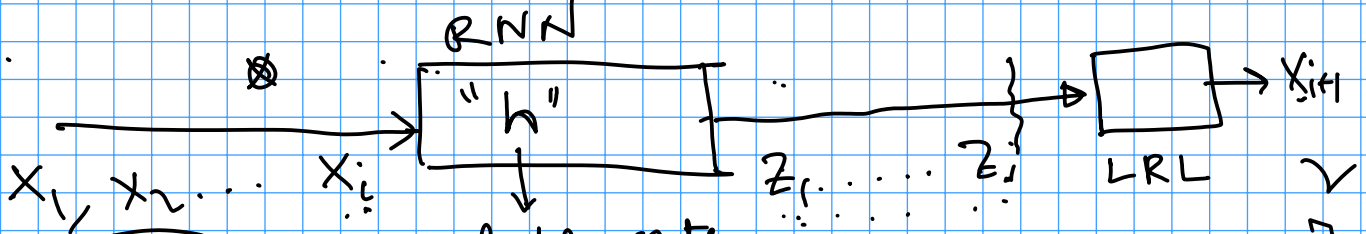
X_{i+1} can depend $X_1 \ldots X_i$ $\quad y_i \to X_i$

(i) stock market prediction
(ii) traffic forecasting
(iii) Dynamical system identificati

$x_1, x_2 \ldots x_i \xrightarrow{?} (x_{i+1}, y_{i+1})$

$y_1\ y_2 \ldots y_i$

Recurrent Neural Network.



hidden state

$h_j = F_1(h_{j-1}, x_j)$

$z_j = F_2(h_j)$

$\left(z_1 - RNN(\ldots)\right)^2 + \left(z_2 - RNN(\ldots)\right)^2 \ldots$

* RNNs can simulate any algorithm (Turing Machine)

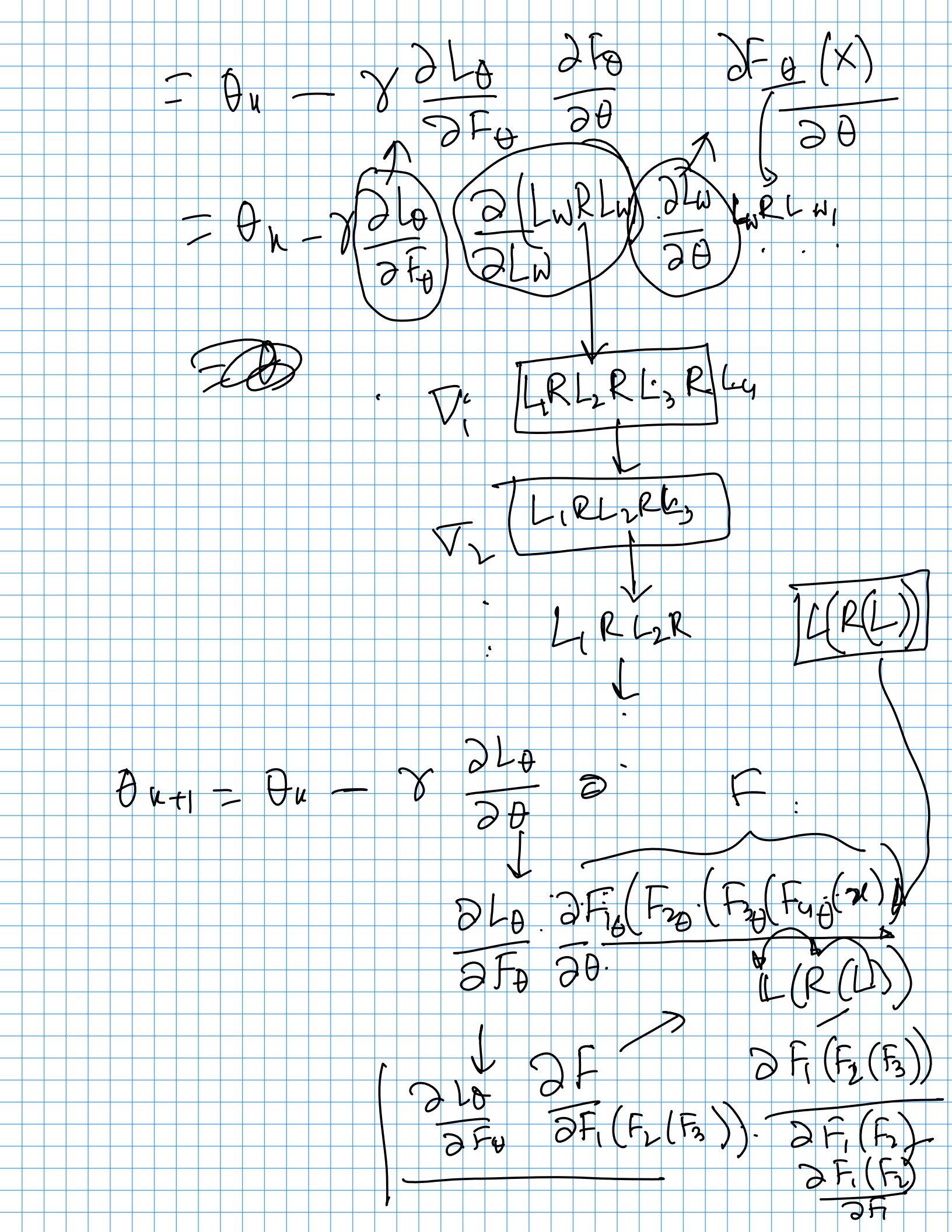
How to optimize the traing objective.

→ Back propagation.

$$\min_\theta \boxed{\left(y - F_\theta(x)\right)^2} \longrightarrow L_\theta(x, y)$$

$$\theta_{k+1} = \theta_k - \gamma \frac{\partial L_\theta(x,y)}{\partial \theta}\Big|_{\theta=\theta_k} \qquad \left(y - F_\theta(x,y)\right)^2$$

$$= \theta_k - \gamma \frac{\partial L_\theta}{\partial F_\theta} \cdot \frac{\partial F_\theta}{\partial \theta}\Big|_{\theta=\theta_k}$$


$= \theta_k - \gamma \frac{\partial L_\theta}{\partial F_\theta} \frac{\partial F_\theta}{\partial \theta}$
$\frac{\partial F_\theta(x)}{\partial \theta}$
$= \theta_k - \gamma \frac{\partial L_\theta}{\partial F_\theta} \frac{\partial (L_W R L_W)}{\partial L_W} \frac{\partial L_W}{\partial \theta}$
$L_W R L_{W_1}$
∇_1
$L_1 R L_2 R L_3 R L_4$
∇_2
$L_1 R L_2 R L_3$
$L_1 R L_2 R$
$L(R(L))$
$\theta_{k+1} = \theta_k - \gamma \frac{\partial L_\theta}{\partial \theta}$
F
$\frac{\partial L_\theta}{\partial F_\theta} \frac{\partial F_{1\theta}(F_{2\theta}(F_{3\theta}(F_{4\theta}(x))))}{\partial \theta}$
$L(R(L))$
$\frac{\partial L_\theta}{\partial F_4} \frac{\partial F}{\partial F_1(F_2(F_3))}$
$\frac{\partial F_1(F_2(F_3))}{\partial F_1(F_2)} \frac{\partial F_1(F_2)}{\partial F_1}$