- deep convnet → gradient vanishing when back prop
- Residual Networks

solve it thru

X (residual → to solve the vanishing problem

F(x)

F(x) + x
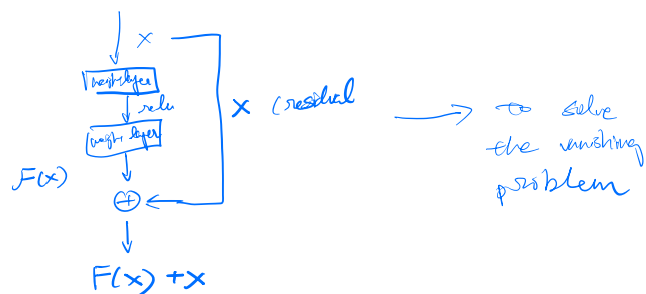
Natural Language Processing :

Glove   Word representation :
- Word Net
- one-hot vectors — { vector size fixed
                       sparsity
                       no natural notion of similarity.

---

word-coding matrix?

PPMI matrix → very sparse.

for sparsity : do SVD

minimize $\| M^{PPMI} - W V^T \|_F^2$
W, V

                                    forbinious
                                    element wise summation

        w×d   w×r   d×r

→ redundant info within the sparse matrix.

SVD : compression

Word2Vec

---

on page 25
difference between two group of words,
if similar relations,
vector looks similar

---

problem: Bank of China v. river bank
                 └ solution: contextual word embedding.

mission: text classification.

GRU $\rightarrow$ simplified
       of
       LSTM