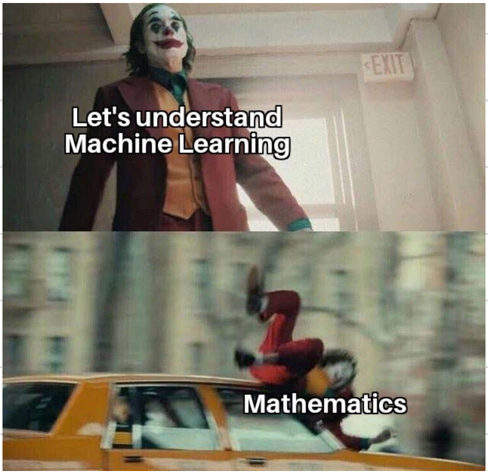
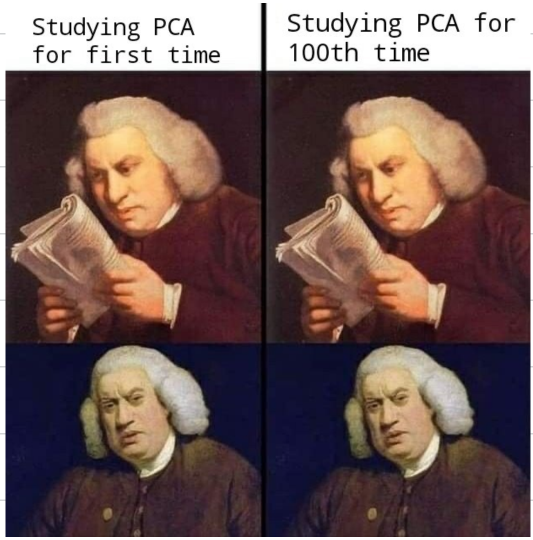
Session-10 # OPTIMIZATION-5 - PCA Feb 17, 2024





AGENDA

1. BATCH - GRADIENT DESCENT
2. PCA- INTUITION

Vanilla Gradient Descent

$$w_{new} = w_{old} - \eta * \sum_{i=1}^{n} \frac{\partial L}{\partial w}$$

n → no of rows

→ Dataset → 1 Billion Rows
Training

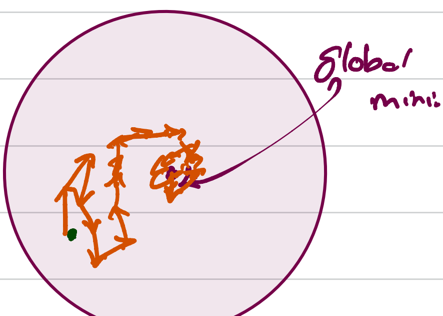
I will have to iterate 1 Billion rows for update weights

→ I take one single row

$$w_{new} = w_{old} - \eta * \sum_{i=1}^{n} \frac{\partial L}{\partial w}$$

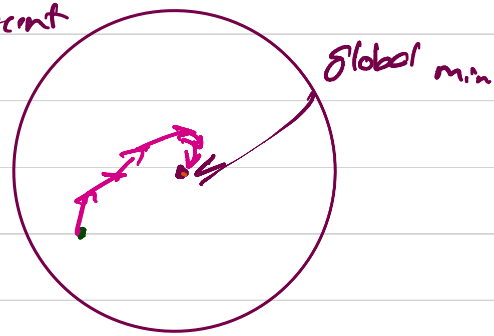
n → no of rows

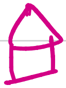
Stochastic gradient descent

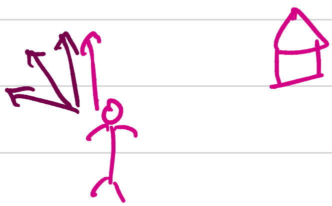


→ For batch gradient descent

When I consider batch of data → updating w

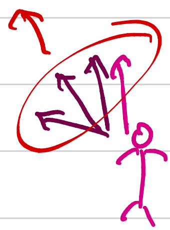




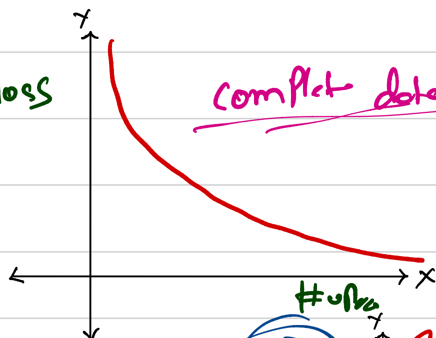


10 → memory

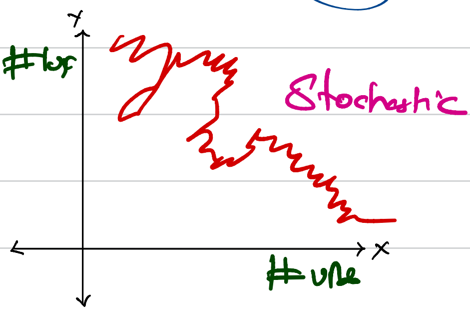
16



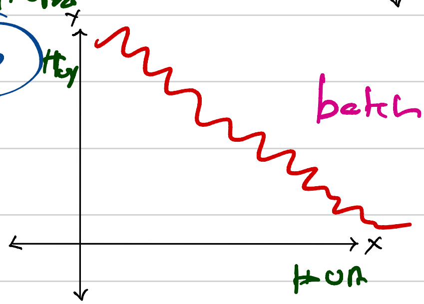
1



2



3



→ $\underbrace{\text{batch-size}}_{\text{not to use mor than } \underline{32}} \propto \frac{1}{\text{training time}}$.

not to use mor than 32 } size of batch

When using $\underline{\geq 32}$ my gradients are too much averaged out

batch-size — 300 $\frac{\partial L}{\partial w}$ avg. 300 rows

→ batch size = 10 / data-points = 100

# of batch = $\frac{100}{10}$ = $\underline{10}$

DIMENSIONALITY REDUCTION

→ 1000 Feature

→ Problems with higher dimensional dataset

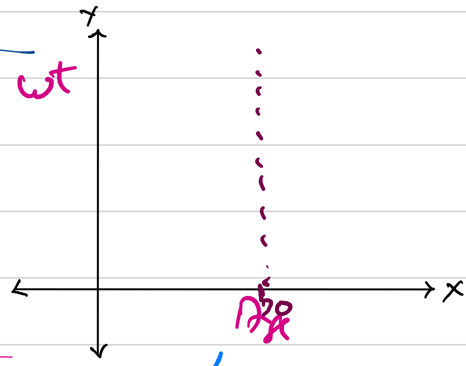
1. Visualization → tough
2. Computationally expensive → training + inference
3. ML models struggle with high dimension data

Price (Open, high, low, close)

→ In PCA → reduce dimensions

↳ without losing much information

Weight	Age	Diabets
20	30	
30	30	
25	30	
45	30	
64	30	

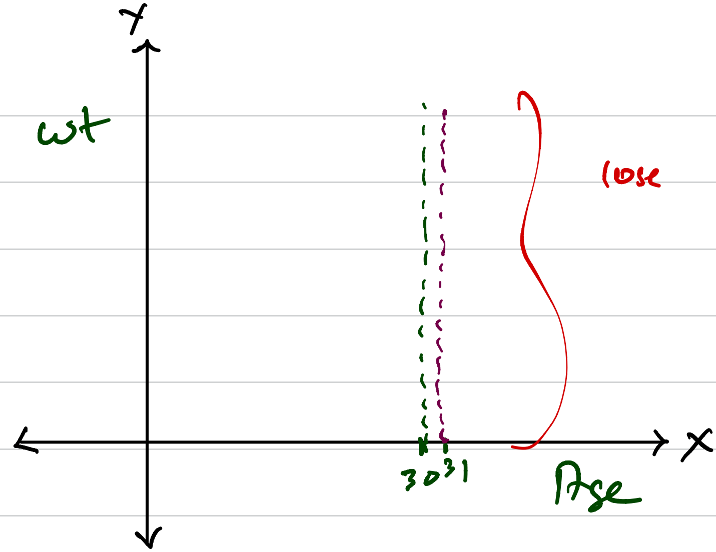


low Variance

→ Feature with high Variance have more Potential to explain TARGET

∝ could be impt

low Varian → def not importa



lose some data

Sacrifice you can make

b.coz

low variance

→ 2nd Situation

PC2

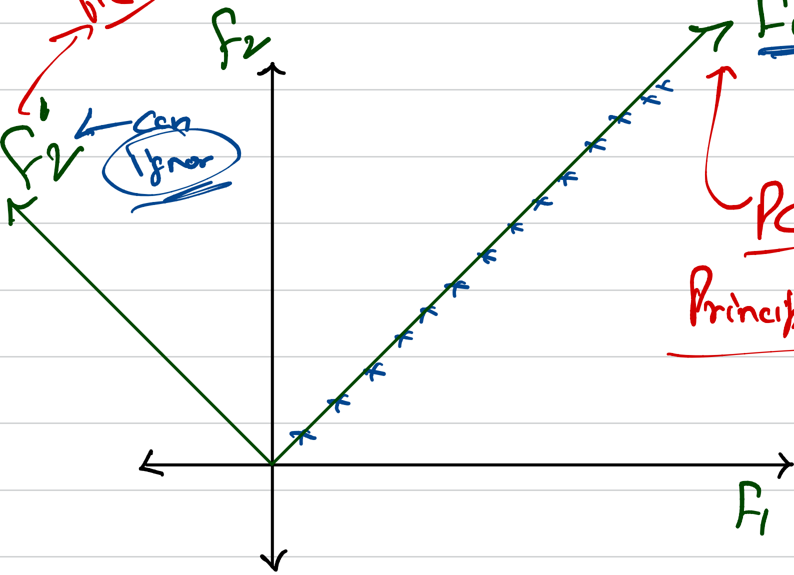
can Ignore

$F_1' = F_1 + 2F_2$

PC1

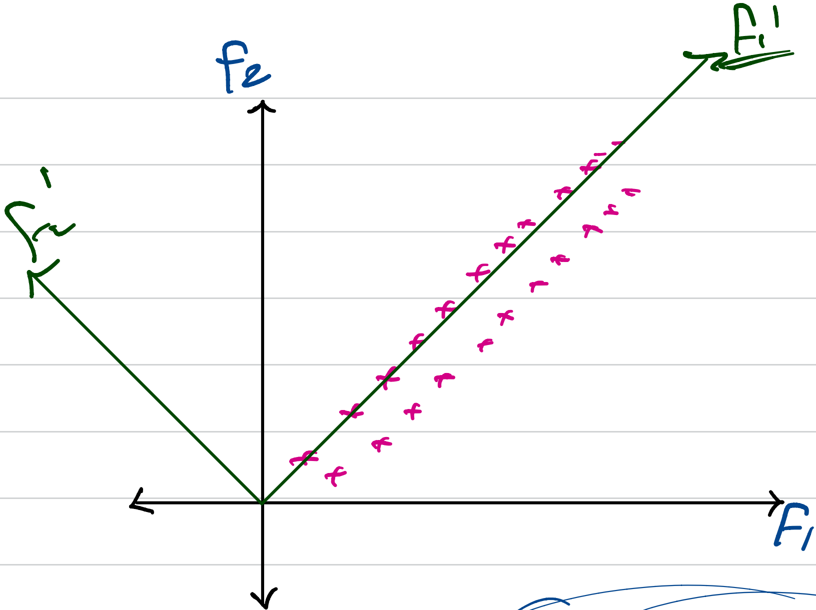
Linearly correlated

Principle component 2

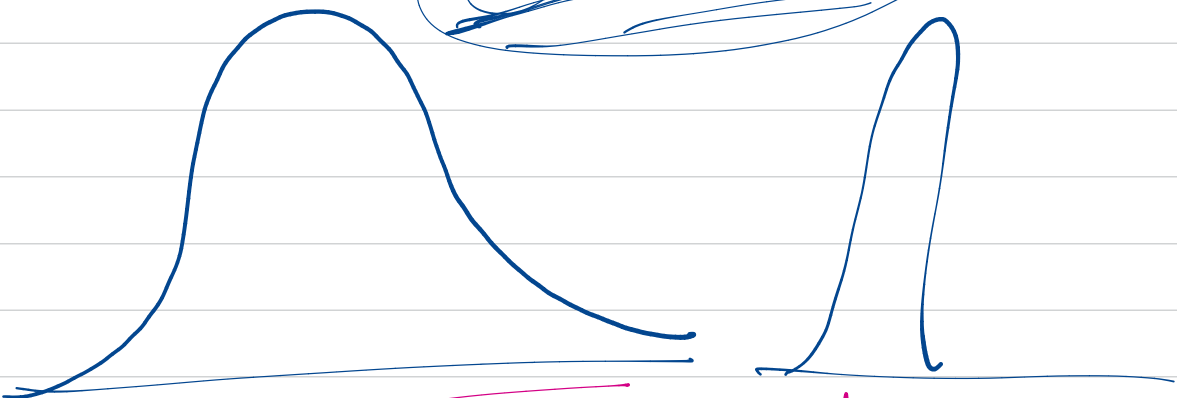


F_1	F_2
1	2
3	4
5	6
7	8

$F_1 + 2F_2$
5
11
17
23



$d' < d$ original one



Wt	Age
2000g	20
5000g	30

Any Problem??

$$(\hat{y}_w - 2000)^2$$

unit makes Problem

mean = 0

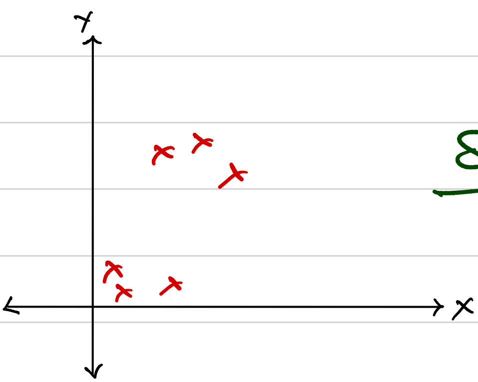
Var = 1

$$[z_1, z_2, \ldots, z_n]$$

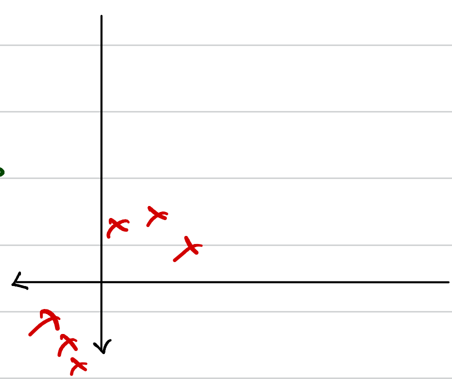
mean = 0

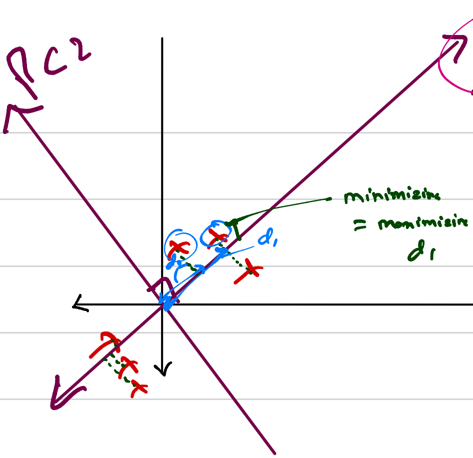
Var = 1 / std = 1

→ Standardization → Primary Pre of PCA



Standardize →



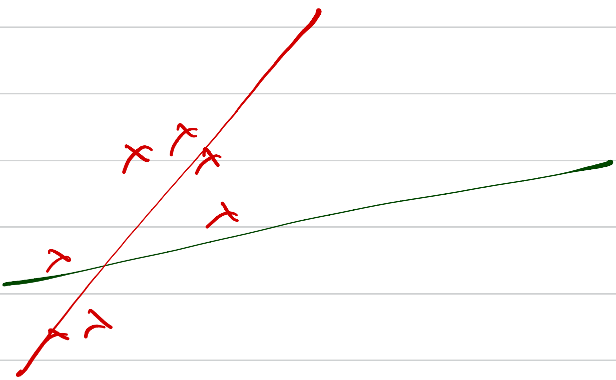


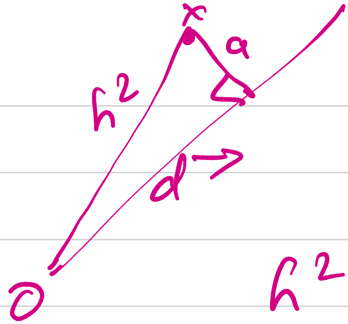
green line dist to be max or min?

$$d_1^2 + d^2 + \cdots d_n^2$$

$$= \frac{\text{sum of square}}{n-1}$$

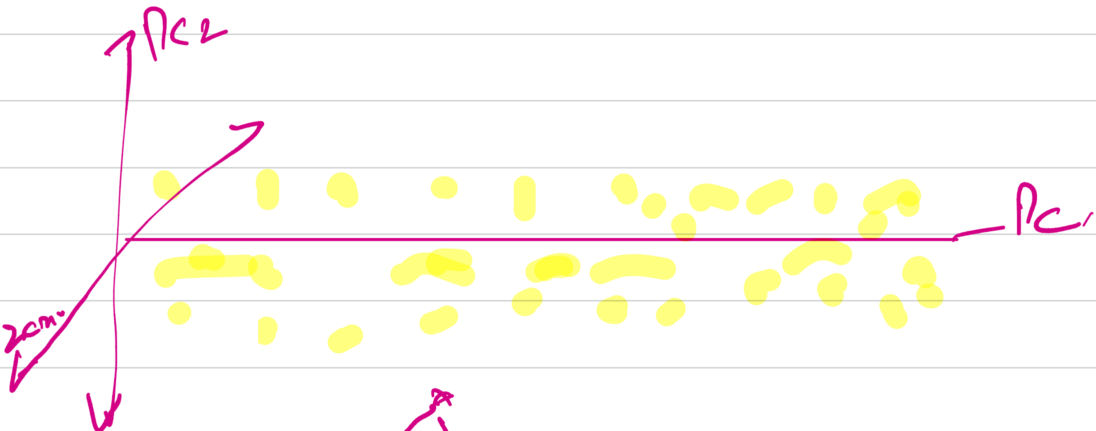
Variance explained by PC1

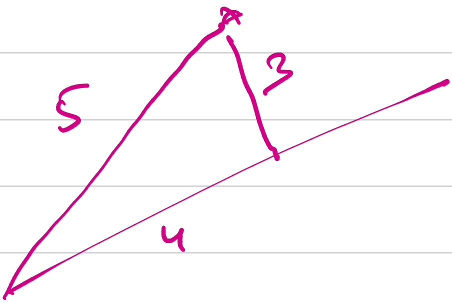




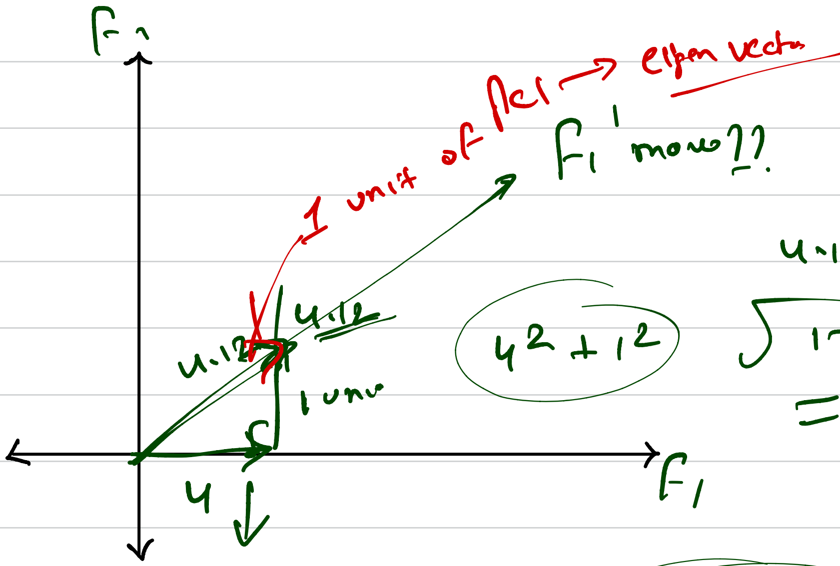
$$h^2 = a^2 + d^2$$

$$a^2 = \underline{h^2 - d^2}$$





$$\underline{5^2 = 3^2 + 4^2}$$



$$4^2 + 1^2$$

$$\sqrt{17} = 4.12$$

$$\frac{4^2}{4.12^2} + \frac{1^2}{4.12^2}$$

$$\frac{4}{4.12}, \quad \frac{1}{4.12}$$

$$\frac{4}{4.12} \rightarrow 0.97 \qquad \frac{1}{4.12} \rightarrow 0.242$$

Mixing 4 units of F_1 & 1 unit of F_2

⇓

PC1 → 4.12 units

Mixing 0.97 of F_1 & 0.242 of F_2 → 1 unit of F_1'

1 units of my PC1 → Eigen Vector

$$\frac{SS(\text{distance of } PC1)}{n-1} = \text{Variance explains by PC1}$$

$$\frac{SS(\text{distance of } PC2)}{n-1} = \text{Variance explains by PC2}$$

d dimen → how to find out d'

$Var(PC1) = 15$ | $15/15+2 = 0.88$

$Var(PC2) = 2$ | $2/15+2 = 0.12$

C_1 | C_2 | C_3 | C_4

PC1 ↓ 60%

PC2 ↓ 35%

PC3 ↓ 7%

PC4 ↓ 2%

PC1 + PC2 % Var = 95%

15 Column → 5, 6

$\underbrace{\phantom{15 \text{ Column}}}$

10 Columns

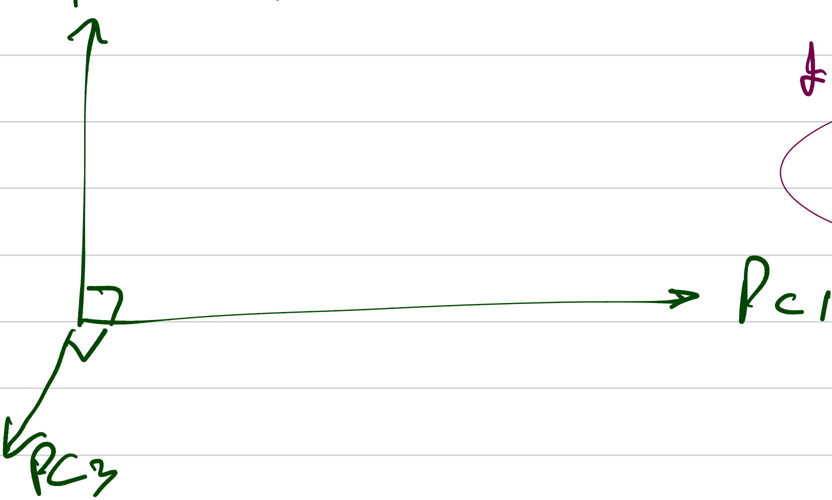
60, 20, 10 — — — —

20 60

10 col → 70 % Varia

Neural Network → Train & Inference

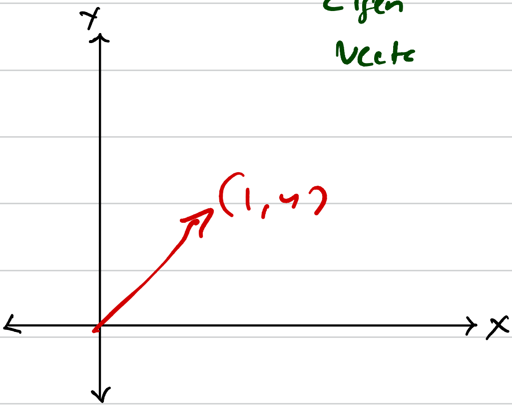
Latency << 0



2D

$$\begin{bmatrix} -6 & 3 \\ 4 & 5 \end{bmatrix}\begin{bmatrix} 1 \\ 4 \end{bmatrix} = \begin{bmatrix} 6 \\ 24 \end{bmatrix} \Rightarrow 6\begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

$2 \times 2 \quad 2\times 1$ Eigen Vecto 2×1 Eigen Value   Ei vecto



$$w = w - \eta \times (w^T x + w_0)\, y_i$$

$$\frac{\partial \left(\cancel{w_1 x_1} + \cancel{w_2 x_2} \cdots \cancel{w_n x_n} + \cancel{w_0} \right) y_i}{\partial \vec{w}} \rightarrow \sum_{i=1}^{n} x_1 \times y_i$$

$$\frac{\partial L}{\partial w_1} = w_1 - \eta \times \sum_{i=1}^{n} y_i \times x_1$$

$$- \eta \times (-1) \times x_1$$

$\times \quad x \times \eta$

$$w_0 - \eta \times (-1) \times w_0$$

$$w_0 + \eta \times 1$$

CV NLP RL