H01 - 1

1 Study Group

a) Yuxiang Gao

Won Tel:15104806995

Weiran_Liu@berkeley.edu

I work on this problem really hard, and some of the problems are too ambiguous (5#) that I change my thoughts multiple times.

b) I certify that all solutions are entirely in my words and I have not looked at another student's solutions. I have credited all external sources in this write up.

HO1-2 Sample Complexity of coupon collecting.

a) [Reference at wikipedia].

Suppose ~~R.V.~~ T is the time all distinct cards are collected.

T can be divided into d R.V.s.

Say $T = \sqcup \sqcup \sqcup \cdots\cdots \underset{}{\overset{t_d}{\sqcup}}$

t_1 being the ~~first~~ time for first card.

t_2: time for collecting second card after the first being collected.

t_3: " third " $1^{st} + 2^{nd}$ "

⋮ etc.

Thus $T = t_1 + t_2 + \cdots + t_d \Rightarrow E[T] = E[t_1] + E[t_2] + \cdots + E[t_d]$

Since $t_i \sim \text{Geometric}(p_i, n)$ where $p_i = \frac{d-i+1}{d}$

and $E[t_i] = \frac{1}{p_i}$

Thus $E[T] = \frac{1}{p_1} + \cdots + \frac{1}{p_d} \approx 225$ (when $d = 50$)

b) the probability of everyday FAIL to win $\left(1 - \frac{1}{d}\right)$

$\mathbb{P}(\text{"fail } n \text{ days"}) = \left(1 - \frac{1}{d}\right)^n$ $\quad \mathbb{P}(\text{"win"}) = 1 - \left(1 - \frac{1}{d}\right)^n \geq 1 - \delta.$

$$n \geq \frac{\log \delta}{\log 1 - \frac{1}{d}}$$

c) $\mathbb{P}(\text{"win in } \alpha d \text{ days"}) = 1 - \left(1 - \frac{1}{d}\right)^{\alpha d}$

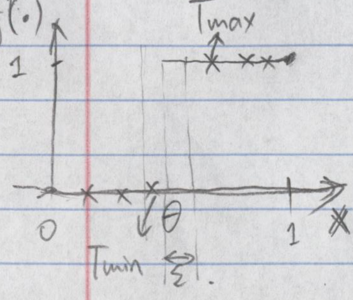
d) $$\lim_{d \to \infty} 1 - \left(1 - \frac{1}{d}\right)^{\alpha d} = 1 - e^{-\alpha}$$

e) $\mathbb{P}(\text{"fails in all } n \text{ samples"}) = \left(1 - \frac{1}{d}\right)^n$

$$1 - \left(1 - \frac{1}{d}\right)^n \geq 1 - \delta$$

$$n \geq \frac{\log \delta}{\log 1 - \frac{1}{d}}$$

H01-3

a) 

$T_{max} - \theta > \varepsilon$ describe the event:

~~all~~ samples fail to locate within $(\theta, \theta+\varepsilon)$

$$\mathbb{P}(T_{max} - \theta > \varepsilon) = \mathbb{P}(\text{"} \downarrow \text{"}) = (1-\varepsilon)^n$$ (n is the number of samples)

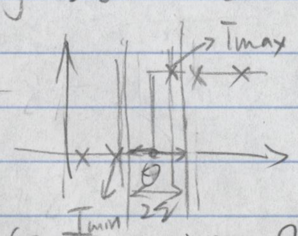
b) for the same reason $\mathbb{P}(\theta - T_{min} > \varepsilon) = (1-\varepsilon)^n$

c) $\mathbb{P}$("Have an estimate for θ within accuracy of 2ε") $\geq 1-\delta$

$\Updownarrow$

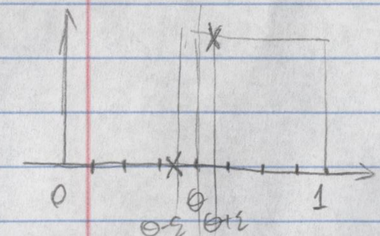
$\mathbb{P}(T_{max} - T_{min} \leq 2\varepsilon) \geq 1-\delta$

$\mathbb{P}(T_{max} - \theta + \theta - T_{min} \leq 2\varepsilon) \geq 1-\delta$ means



$\mathbb{P}(T_{max} - \theta + \theta - T_{min} \leq 2\varepsilon) \leq \mathbb{P}(T_{max} - \theta \leq \varepsilon) + \mathbb{P}(\theta - T_{min} \leq \varepsilon) \geq 1-\delta$

$$2\left(1-(1-\varepsilon)^n\right) \geq 1-\delta \Rightarrow n \geq \frac{\log\frac{1+\delta}{2}}{\log 1-\varepsilon}$$

d) 

I will sample the points evenly.

$\mathbb{P}$("Achieve ε-close") $= \mathbb{P}(T_{max} - T_{min} \leq 2\varepsilon)$

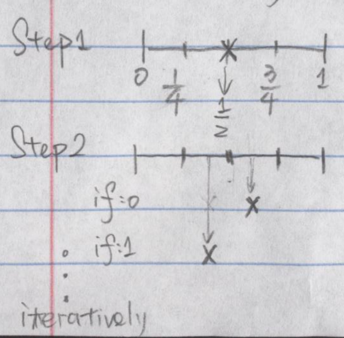
In this way, $\frac{T_{min}+T_{max}}{2}$ will lies in $(\theta-\varepsilon, \theta+\varepsilon)$.

$\ell = T_{max} - T_{min} \leq 2\varepsilon$

$$\mathbb{P}(\ell \leq 2\varepsilon) = \frac{2\varepsilon}{\ell} = \frac{2\varepsilon}{\frac{1}{n}} \geq 1-\delta$$

$$n \geq \frac{1-\delta}{2\varepsilon}$$

e) Use Dichotomy to sample points



$$\mathbb{P}(\text{"one point in } \varepsilon\text{-close interval"}) = \frac{2\varepsilon}{(\frac{1}{2})^n} \geq 1-\delta$$

$$n \geq \frac{\log\frac{1-\delta}{2\varepsilon}}{\log 2}$$

H01 - 4

a) Right EigenVector / EigenValues | Left EigenVector / EigenValues

A $\begin{bmatrix} \frac{\sqrt{6}}{3} \\ \frac{\sqrt{3}}{3} \end{bmatrix} \begin{bmatrix} -\frac{\sqrt{6}}{3} \\ \frac{\sqrt{3}}{3} \end{bmatrix}$ $3+\sqrt{2},\ 3-\sqrt{2}$ | $\begin{bmatrix} \frac{\sqrt{3}}{3} \\ \frac{\sqrt{6}}{3} \end{bmatrix} \begin{bmatrix} -\frac{\sqrt{3}}{3} \\ \frac{\sqrt{6}}{3} \end{bmatrix}$ $3+\sqrt{2},\ 3-\sqrt{2}$

B $\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$ $7,\ 3$ | $\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{bmatrix}$ $7,\ 3$

$A^2 = \begin{bmatrix} 11 & 12 \\ 6 & 11 \end{bmatrix}$ $\begin{bmatrix} \frac{\sqrt{6}}{3} \\ \frac{\sqrt{3}}{3} \end{bmatrix} \begin{bmatrix} -\frac{\sqrt{6}}{3} \\ \frac{\sqrt{3}}{3} \end{bmatrix}$ $11+6\sqrt{2},\ 11-6\sqrt{2}$ | $\begin{bmatrix} \frac{\sqrt{3}}{3} \\ \frac{\sqrt{6}}{3} \end{bmatrix} \begin{bmatrix} -\frac{\sqrt{3}}{3} \\ \frac{\sqrt{6}}{3} \end{bmatrix}$ $11+6\sqrt{2},\ 11-6\sqrt{2}$

$B^2 = \begin{bmatrix} 29 & 20 \\ 20 & 29 \end{bmatrix}$ $\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{bmatrix}$ $49\ \ 9$ | $\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$ $49,\ 9$

$AB = \begin{bmatrix} 19 & 16 \\ 11 & 17 \end{bmatrix}$ $\begin{bmatrix} 0.793 \\ 0.610 \end{bmatrix} \begin{bmatrix} -0.746 \\ 0.666 \end{bmatrix}$ $\overset{31.30}{18+\sqrt{177}},\ \overset{4.70}{18-\sqrt{177}}$ | $\begin{bmatrix} 0.666 \\ 0.746 \end{bmatrix} \begin{bmatrix} -0.610 \\ 0.793 \end{bmatrix}$ $18+\sqrt{177},\ 18-\sqrt{177}$

$BA = \begin{bmatrix} 17 & 16 \\ 11 & 19 \end{bmatrix}$ $\begin{bmatrix} -0.746 \\ 0.666 \end{bmatrix} \begin{bmatrix} -0.793 \\ 0.610 \end{bmatrix}$ $\overset{31.30}{18+\sqrt{177}},\ \overset{4.70}{18-\sqrt{177}}$ | $\begin{bmatrix} -0.610 \\ -0.793 \end{bmatrix} \begin{bmatrix} -0.666 \\ 0.746 \end{bmatrix}$ $18+\sqrt{177},\ 18-\sqrt{177}$

b)

$$A = \begin{bmatrix} -0.763 & -0.646 \\ -0.646 & 0.763 \end{bmatrix} \begin{bmatrix} 4.54 & 0 \\ 0 & 1.54 \end{bmatrix} \begin{bmatrix} -0.646 & -0.763 \\ -0.763 & 0.646 \end{bmatrix} \quad AB = \begin{bmatrix} -0.78 & -0.63 \\ -0.63 & 0.78 \end{bmatrix} \begin{bmatrix} 31.71 & 0 \\ 0 & 4.64 \end{bmatrix} \begin{bmatrix} -0.68 & -0.73 \\ -0.73 & 0.68 \end{bmatrix}$$

$$B = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 7 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \quad BA = \begin{bmatrix} -0.73 & -0.68 \\ -0.68 & 0.73 \end{bmatrix} \begin{bmatrix} 31.71 & 0 \\ 0 & 4.64 \end{bmatrix} \begin{bmatrix} -0.63 & -0.78 \\ -0.78 & 0.63 \end{bmatrix}$$

$$A^2 = \begin{bmatrix} -0.794 & -0.607 \\ -0.607 & 0.794 \end{bmatrix} \begin{bmatrix} 20.402 & 0 \\ 0 & 2.402 \end{bmatrix} \begin{bmatrix} -0.607 & -0.794 \\ -0.794 & +0.607 \end{bmatrix}$$

$$B^2 = U_B D_B^2 U_B^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 49 & 0 \\ 0 & 9 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

$$C = \overset{4\times 2}{\begin{bmatrix} \vec{B} \\ A \end{bmatrix}} = \begin{bmatrix} -4.22 & 2.36 & -8.50 & -2.10 \\ -3.43 & -3.96 & 2.60 & 8.11 \\ -5.87 & 6.64 & 4.58 & 7.12 \\ -6.00 & -5.89 & 6.84 & 5.42 \end{bmatrix} \begin{bmatrix} 8.34 & 0 \\ 0 & 3.39 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -0.69 & +0.73 \\ -0.73 & -0.69 \end{bmatrix}$$

c) $(\lambda - A_{ii})\, x_i = \sum_{j \neq i} A_{ij} x_j \iff \vec{x}$ is right eigen vector.

rhs→lhs: for a fixed i $\quad \lambda x_i = \sum_{j=1}^{n} A_{ij} x_i$

Then $\lambda \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \vec{A_1} \\ \vdots \\ \vec{A_n} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ $\quad \therefore \vec{x}$ is right eigenvector

lhs→rhs: if $\lambda \vec{x} = A \vec{x}$

$$\lambda \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \vec{A_1} \\ \vdots \\ \vec{A_n} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

for ith row:

$$\lambda x_i = \vec{A_i} \vec{x}$$

$$\lambda x_i = \sum A_{ij} x_j$$

$$\lambda x_i - A_{ii} x_i = (\lambda - A_{ii})\, x_i = \sum_{j \neq i} A_{ij} x_j$$

d) $(\lambda - A_{ii})\, x_i = \sum_{j \neq i} A_{ij} x_j$

$$|\lambda - A_{ii}|\,|x_i| = \left|\sum_{j \neq i} A_{ij} x_i\right| \leq \sum_{j \neq i} |A_{ij} x_j| = \sum_{j \neq i} |A_{ij}|\,|x_j| \leq \sum_{j \neq i} |A_{ij}|\,|x_i|$$

$$\therefore |\lambda - A_{ii}| \leq \sum_{j \neq i} |A_{ij}|$$

H01-5

a)

$$\begin{array}{c|c} & \begin{array}{ccccc} \vec{t}_1 & \vec{t}_2 & \vec{t}_3 & \cdots & \vec{t}_d \end{array} \\ \begin{array}{c} a^{(1)} \\ a^{(2)} \\ \vdots \\ a^{(n)} \end{array} & \begin{bmatrix} a_1^{(1)} & a_2^{(1)} & a_3^{(1)} & & a_d^{(1)} \\ a_1^{(2)} & a_2^{(2)} & a_3^{(2)} & \cdots & a_d^{(2)} \\ \vdots & & & & \\ a_1^{(n)} & a_2^{(n)} & a_3^{(n)} & & a_d^{(n)} \end{bmatrix} \end{array} \qquad \begin{array}{c} \vec{b} \\ \begin{bmatrix} b^{(1)} \\ b^{(2)} \\ \vdots \\ b^{(n)} \end{bmatrix} \end{array}$$

sample points

the sequence of linear predictor

① first use column $\vec{t}_1$ estimate $\vec{b}$ using OLS obtain w_{11}

② " $\vec{t}_1, \vec{t}_2$ estimate $\vec{b}$ " $w_{21}\ w_{22}$

⋮

ⓓ use $\vec{t}_1 \cdots \vec{t}_d$ estimate $\vec{b}$ " $w_{d1}\ w_{d2} \cdots w_{dd}$

These are the sequence of Linear Predictor

b)

Step 1: use $\begin{bmatrix} a_1^{(1)} & a_2^{(1)} & & a_i^{(1)} \\ a_1^{(2)} & a_2^{(2)} & \cdots & a_i^{(2)} \\ \vdots & \vdots & & \\ a_1^{(n)} & a_2^{(n)} & & a_i^{(n)} \end{bmatrix}$ to estimate $\begin{bmatrix} a_{i+1}^{(1)} \\ a_{i+1}^{(2)} \\ \vdots \\ a_{i+1}^{(n)} \end{bmatrix}$ using OLS

n sample points

Step 2: Say the estimated $\begin{bmatrix} a_{i+1}^{(1)} \\ a_{i+1}^{(2)} \\ \vdots \\ a_{i+1}^{(n)} \end{bmatrix}$ is $\begin{bmatrix} \hat{a}_{i+1}^{(1)} \\ \hat{a}_{i+1}^{(2)} \\ \vdots \\ \hat{a}_{i+1}^{(n)} \end{bmatrix}$

compute $\hat{\varepsilon}_{i+1} = \begin{bmatrix} \hat{a}_{i+1}^{(1)} \\ \hat{a}_{i+1}^{(2)} \\ \vdots \\ \hat{a}_{i+1}^{(n)} \end{bmatrix} - \begin{bmatrix} a_{i+1}^{(1)} \\ a_{i+1}^{(2)} \\ \vdots \\ a_{i+1}^{(n)} \end{bmatrix}$

Step 3: use $\hat{\varepsilon}_{i+1}$ and $\hat{b}_i$ to estimate $\vec{b}$

Thus $\hat{b}_{i+1}$ is the linear Combination of $\hat{\varepsilon}_{i+1}$ and $\hat{b}_i$

Doing this Iteratively

c) Not Equivalent.

The two approaches both give out a sequence of Predictor.

As time step goes, both predictor becomes better.

However the second one is better than the first one.

The first one is to project on to the linear space of the sample points. Because time sequence has strong correlation (dependency), this means the feature is highly dependent. The dependency made the model rather dificult to fit, since one single feature could be the linear combination of many previous features. Thus, tune one weight may affect the model in different aspects.

And what the second way do is to eliminate the dependency. By computing the error, $\hat{\varepsilon}_{i+1} = \hat{\vec{a}}_{i+1} - \vec{a}_{i+1}$, it gives us the new information provided by $\vec{a}_{i+1}$. This is because:

$\hat{a}_{i+1}$ is the projection of $\vec{a}_{i+1}$ on the Linear space $[a_1 \cdots a_i]$

Thus, $\hat{\varepsilon}_{i+1} \perp$ the previous "information feature vector".

$\hat{\vec{b}}_i$ is the estimate of $\vec{b}$ on the ith step, It contains all the information in the first i time step, because it lies in first i feature Linear Space. (projection)

Thus, orthogonal to $\vec{a}_{i+1}$.

To sum up, what the second approach do is to break down the dependency and add new information step by step.

Thus, the second is way much better than the first one.

d)

Using Frobenius normal is equivalent to put element in X into a long long Vector them minimize its 2 norm.

e) the first is Frobenius Norm and second is 2-norm. They write in different form but equivalent.

HO1-6 a), b) code.

a)

$$\begin{bmatrix} x^{(1)}(t+1) \\ \vdots \\ x^{(n)}(t+1) \end{bmatrix} = \begin{bmatrix} x^{(1)}(t) \\ \vdots \\ x^{(n)}(t) \end{bmatrix} \omega_1 + \begin{bmatrix} u^{(1)}(t) \\ \vdots \\ u^{(n)}(t) \end{bmatrix} \omega_2$$

$$\Downarrow$$

$$\underbrace{\begin{bmatrix} x^{(1)}(t+1) \\ \vdots \\ x^{(n)}(t+1) \end{bmatrix}}_{\vec{b}} = \underbrace{\begin{bmatrix} x^{(1)}(t) & u^{(1)}(t) \\ \vdots & \\ x^{(n)}(t) & u^{(n)}(t) \end{bmatrix}}_{A} \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix}$$

$$\begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} = (A^T A)^{-1} A^T b$$

b)

$$\begin{bmatrix} X_{t+1} \end{bmatrix} = A \begin{bmatrix} X_t \end{bmatrix} + B \begin{bmatrix} U_t \end{bmatrix}$$

array ay

$$\underset{3\times n}{X_{t+1}} = \underset{3\times 6}{[A\ B]} \underset{6\times n}{\begin{bmatrix} X_t \\ U_t \end{bmatrix}}$$

$$\text{minimize} \left\| X_{t+1} - [A\ B] \begin{bmatrix} X_t \\ U_t \end{bmatrix} \right\|_F$$

X_{t+1}

$$\begin{bmatrix} x_{1}, x_{2}, x_{3}, u_1, u_2, u_3 \\ \\ \\ \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \\ b_{11} \\ b_{12} \\ b_{13} \\ \vdots \end{bmatrix}$$

6 — 18

HO1 - b (c)

$$\begin{cases} y_1 = x_i \\ y_2 = \dot{y}_1 - h_1 x_{i-1} = \dot{x}_i - h_1 x_{i-1} \end{cases}$$

$$\begin{cases} \dot{y}_2 = \ddot{x}_i - h_1 \dot{x}_{i-1} = \alpha y_1 + \beta y_2 + h_2 x_{i-1} \\ \ddot{x}_i - h_1 \dot{x}_{i-1} = \alpha x_i + \beta(\dot{x}_i - h_1 x_{i-1}) + h_2 x_{i-1} \\ \ddot{x}_i = a x_i + b \dot{x}_i + c x_{i-1} + d \dot{x}_{i-1} \end{cases}$$

$\alpha = a;\ \beta = b;\ \beta h_1 + h_2 = c;\ h_2 = d \qquad h_1 = \frac{c-h_2}{\beta} = \frac{c-d}{b}$

choose state var:

$$\begin{cases} y_1 = x_i \\ y_2 = \dot{y}_1 - \frac{c-d}{b} x_{i-1} \end{cases} \qquad \begin{cases} y_2 = \dot{y}_1 - \frac{c-d}{b} x_{i-1} \Rightarrow \dot{y}_1 = y_2 + \frac{c-d}{b} x_{i-1} \\ \dot{y}_2 = \cancel{ax_i}\ a y_1 + b y_2 + d x_{i-1} \end{cases}$$

$$\begin{bmatrix} \dot{y}_1 \\ \dot{y}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ a & b \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + \begin{bmatrix} \frac{c-d}{b} \\ d \end{bmatrix} x_{i-1} + w(t) \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$\cancel{x_i}$

test: $(\dot{y}_2 + \frac{c-d}{b} \dot{x}_{i-1})$

Training process: ① generate $\dot{y}_1 = \dot{x}_i$, $\dot{y}_2 = \ddot{x}_i - \boxed{\frac{c-d}{b}} \dot{x}_{i-1}$

$y_1 = x_i$, $y_2 = \dot{x}_i - \boxed{\frac{c-d}{b}} x_{i-1}$

(d) $\min \| \ddot{x}_i - \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}^T \begin{bmatrix} x_i \\ \dot{x}_i \\ x_{i-1} \\ \dot{x}_{i-1} \end{bmatrix} \|_2^2 \Rightarrow \min_{\vec{v}} \| b - \vec{v} A \|_2$ $\quad 1\times n \quad 4\times$

$\vec{v} = [a, b, c, d]^T$

→ 1×n row vector

4×n row vector

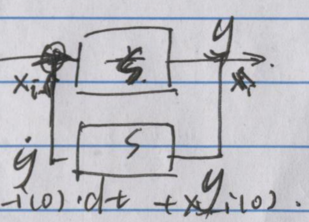
n samples

$\vec{v} = (A \cdot A^T)^{-1} A \cdot b^T$

Question:

$\begin{cases} \text{Discrete} \\ \text{Continuous.} \end{cases}$

(e) $\begin{bmatrix} \boxed{\begin{matrix} x\_i(0) \\ dot\_x\_i(0) \end{matrix}} & \cdots & \begin{matrix} x\_i(150) \\ dot\_x\_i(150) \end{matrix} \end{bmatrix}$



* In our case, car acc ~~is~~ is generated from (p0, v.) continuous variables. However, we sample ~~that~~ our data by f=15Hz. Is this continuous model works. Why don't we use discrete model?

$\frac{x\_i(1) - x\_i(0)}{dt} = dot\_x\_i(0) \Rightarrow x\_i(1) = dot\_x\_i(0) \cdot dt + x\_i(0)$.

Programming

① Data Processing → a. velocity } + limit (xi / xi-1)

b. ~~xi~~ para training (OLS)

② generate $x\_i(t) = \underline{x\_i(t-1)}$

$$\begin{cases} x\_i(t) = x\_i(t-1) + dot\_x\_i(t-1) \frac{1}{15} \\ \ddot{x}\_i(t-1) \\ dot\_x\_i(t) = dot\_x\_i(t-1) + \ddot{x}\_i(t-1) \frac{1}{15} \end{cases}$$

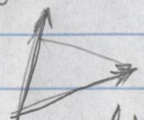
$\ddot{x}\_i \leftarrow x\_i(0)\ \ dot\_x\_i(0)$

$dot\_x\_i(1)$

$xp^{(0)}\ xdp^{(0)}$
$xp^{(1)}\ xdp^{(1)}$

H01-7

In Problem 05, Can we use only the error $\hat{\varepsilon}_i$ to predict $\vec{b}$ every time
Then add those estimate together?



Not as good as the previous method

$w_1 \vec{\varepsilon}_1$

$$w_1(\ldots a_1) + w_2(\hat{\vec{a}}_2 - \vec{a}_2) + \cdots + w_d(\hat{\vec{a}}_d - a_d) = \hat{b}$$

$\perp[a_1\ a_2]$ $\quad$ $\perp[a_1\ a_2\ a_3 \cdots a_{d-1}]$

Although ~~Although~~ the error orthogonal to the previous measurement
not all error orthogonal to each other, say $\hat{\varepsilon}_2$, ε_4
There's still dependency inside the feature.