

Theoretical Machine Learning

Lectured by [Zhihua Zhang](#)

L^AT_EXed by [Chengxin Gong](#)

March 5, 2024

Contents

1	简介	2
2	统计决策理论	2
3	马尔可夫决策过程	3

1 简介

- 机器学习的主要任务: 生成、预测、决策. 生成: $X_1, \dots, X_n \sim F$, 推断分析 F , 无监督学习, GAN, GPT, \dots . 预测: 数据对 $(X^{(1)}, Y^{(1)}), \dots, (X^{(n)}, Y^{(n)})$, $X^{(i)} \in \mathbb{R}^d$ 输入变量, $f: \mathcal{X} \rightarrow \mathcal{Y}, x \in \mathcal{X}, y \in \mathcal{Y}$, 归因, 有监督学习. 决策: 强化学习, Agent \leftarrow action, state, reward \rightarrow 环境.
- 求解问题的途径: 参数/非参数, 频率 (MLE)/贝叶斯.
- 误差模型: 有监督: $X = (X_1, \dots, X_d)^T \in \mathbb{R}^d$, 回归: $Y \in \mathbb{R}$; 分类: $Y \in \{0, 1\}(\{-1, 1\}, \{1, \dots, M\}, \{0, 1\}^M)$; X 随机, Random design(生成模型), $Y = g(X) + \varepsilon \stackrel{\text{or}}{=} g(X, Z), Y^{(i)} = g(X^{(i)}, Z^{(i)})$; X 固定 $X = x$, Fixed design(判别模型), $Y^{(i)} = g(x^{(i)}, Z^{(i)})$. 无监督: $X = g(Z)$ (因子模型: $X = AZ + \varepsilon, Z \in \mathcal{N}(0, 1), \varepsilon \sim \mathcal{N}(0, \Sigma)$).

2 统计决策理论

- Consider a state space Ω , data space \mathcal{D} , model $\mathcal{P} = \{p(\theta, x)\}$, action space \mathcal{A} . Loss function: $\mathcal{L}: \Omega \times \mathcal{A} \rightarrow [-\infty, +\infty]$, measurable, nonnegative. A measurable function $\delta: \mathcal{D} \rightarrow \mathcal{A}$ is called a nonrandomized decision rule. Risk function is defined as $\mathcal{R}(\theta, \delta) = \int \mathcal{L}(\theta, \delta(x)) dP_\theta(x) = \mathbb{E}_\theta \mathcal{L}(\theta, \delta(X))$. Randomized decision: for each $X = x$, $\delta(x)$ is a probability distribution: $[A|X = x] \sim \delta_x$. Risk function for δ : $\mathcal{R}(\theta, \delta) = \mathbb{E}_\theta \mathcal{L}(\theta, A) = \mathbb{E}_\theta \mathbb{E}_a \mathcal{L}(\theta, A|X) = \iint \mathcal{L}(\theta, a) d\delta_x(a) dP_\theta(x)$.
- Example [参数估计]: $\theta \in \Omega, \mathcal{A} = \Omega, \mathcal{L}(\theta, a) = \|\theta - a\|_2^2 \stackrel{\text{or}}{=} \|\theta - a\|_p^p (p \geq 1) \stackrel{\text{or}}{=} \int \log \frac{P_\theta(x)}{P_a(x)} P_\theta(x) dm(x) (\text{KL})$. $\mathcal{R} = \text{Var}(a) + \text{bias}^2(a)$. Bregmass loss: $\phi: \mathbb{R}^d \rightarrow \mathbb{R}$ describe any strictly convex differentiable function. Then $\mathcal{L}_\phi(\theta, a) = \phi(a) - \phi(\theta) - (\phi - a)^T \nabla \phi(a)$.
- Example [Testing]: $\mathcal{A} = \{0, 1\}$ with action “0” associated with accepting $H_0: \theta \in \Omega_0$ and “1”: $H_1: \theta \in \Omega_1$. δ_x is a Bernolli distribution. $\mathcal{L}(\theta, a) = I\{a = 1, \theta \in \Omega_0\} + I\{a = 0, \theta \in \Omega_1\}$. Risk $\mathcal{R}(\theta, \delta) = \mathbb{P}_\theta(A = 1)1_{\theta \in \Omega_0} + \mathbb{P}_\theta(A = 0)1_{\theta \in \Omega_1}$.
- A decision rule δ is called inadmissible if a competing rule δ^* such that $\mathcal{R}(\theta, \delta^*) \leq \mathcal{R}(\theta, \delta)$ for all $\theta \in \Omega$ and $\mathcal{R}(\theta, \delta^*) < \mathcal{R}(\theta, \delta)$ for at least one $\theta \in \Omega$. Otherwise, δ is admissible.
- The maximum risk $\bar{\mathcal{R}}(\delta) = \sup_{\theta \in \Omega} \mathcal{R}(\theta, \delta)$ and the Bayes risk $r(\Lambda, \delta) = \int \mathcal{R}(\theta, \delta) d\Lambda(\theta) = \int \mathcal{L}(\theta, \delta) d\mathbb{P}(x, \theta)$ ($\Lambda(\theta)$ is a prior). A decision rule that minimizes the Bayes risk is called a Bayes rule, that is, $\hat{\delta}: r(\Lambda, \hat{\delta}) = \inf_\delta r(\Lambda, \delta)$. Minimax rule $\delta^*: \sup_{\theta \in \Omega} \mathcal{R}(\theta, \delta^*) = \inf_\delta \sup_{\theta \in \Omega} \mathcal{R}(\theta, \delta)$.
- If risk functions for all decision rules are continuous in θ , if δ is Bayesian for Λ and has finite integrated risk $r(\Lambda, \delta) < \infty$, and if the support of Λ is the whole state space Ω , then δ is admissible.
- $p(\theta|x) = \frac{p_\theta(x)\lambda(\theta)}{\int p_\theta(x)\lambda(\theta)d\theta} := \frac{p_\theta(x)\lambda(\theta)}{m(x)}$. Define the posterior risk of δ : $r(\delta|X = x) = \int \mathcal{L}(\theta, \delta(x)) d\mathbb{P}(\theta|x)$. The Bayes risk $r(\Lambda, \delta)$ satisfies that $r(\Lambda, \delta) = \int r(\delta|x) dM(x)$. Let $\hat{\delta}(x)$ be the value of δ that minimizes $r(\delta|x)$. Then $\hat{\delta}$ is the Bayes rule.
- Application to supervised learning. Case 1: Regression. $(X, Y) \in \mathcal{X} \times \mathcal{Y}, f: \mathcal{X} \rightarrow \mathcal{Y}, \mathcal{A} = \Omega = \mathcal{Y}, \mathcal{D} = \mathcal{X}, \delta = f, \mathcal{L}(Y, f(X)) = \|Y - f(X)\|_p^p, p \geq 1$, risk $R_f = \iint \mathcal{L}(y, f(x)) d\mathbb{P}(x, y) = \mathbb{E}[\mathcal{L}(Y, f(X))] = \mathbb{E}[\mathbb{E}\mathcal{L}(Y, f(X))|X]$. When $p = 2$, $r(f|X = x) = \int \mathcal{L}(y, f(x)) d\mathbb{P}(y|x) = \int |y - f(x)|^2 d\mathbb{P}(y|x)$. 回归函数 $g(x) := \int y d\mathbb{P}(y|x) \Rightarrow R_f = \mathbb{E}|Y - f(X)|^2 = \mathbb{E}|Y - g(X) + g(X) - f(X)|^2 = \mathbb{E}|Y - g(X)|^2 + \mathbb{E}|g(X) - f(X)|^2 \geq \mathbb{E}|Y - g(X)|^2$.
- Case 2: Pattern classification. $Y \in \{0, 1\}, p_0 = P(Y = 0), p_1 = P(Y = 1) = 1 - p_0, \mathbb{E}[\mathcal{L}(Y, f(X))] = P(Y \neq f(X))$. The Bayesian rule (predictor) is given by $f(x) = 1\{\mathbb{P}(Y = 1|X = x) \geq \frac{\mathcal{L}(1,0) - \mathcal{L}(0,0)}{\mathcal{L}(0,1) - \mathcal{L}(1,1)} \mathbb{P}(Y = 0|X = x)\}$. (Proof: $\mathbb{E}[\mathcal{L}(Y, f(X))|X = x] = \begin{cases} \mathbb{E}[\mathcal{L}(Y, 0)|X = x] = \mathcal{L}(0,0)\mathbb{P}(Y = 0|X = x) + \mathcal{L}(1,0)\mathbb{P}(Y = 1|X = x) \\ \mathbb{E}[\mathcal{L}(Y, 1)|X = x] = \mathcal{L}(0,1)\mathbb{P}(Y = 0|X = x) + \mathcal{L}(1,1)\mathbb{P}(Y = 1|X = x) \end{cases}$, 比较大小)
- 连续化: $\mathbb{P}(Y = 1|X = x) = \mathbb{E}(Y|X = x) := g(x)$ (回归), $f(x) = 1\{g(x) \geq \frac{1}{2}\}$. Then $0 \leq \mathbb{P}(\hat{f}(X) \neq Y) - \mathbb{P}(f(X) \neq Y) \leq 2 \int_{\mathcal{X}} |\hat{g}(x) - g(x)| \mu(dx) \leq 2(\int_{\mathcal{X}} |\hat{g}(x) - g(x)|^2 \mu(dx))^{\frac{1}{2}}$.

- 回到 Case 2. $f(x) = 1\{\frac{p(x|y=1)}{p(x|y=0)} \geq \frac{p_0(\mathcal{L}(0,1)-\mathcal{L}(0,0))}{p_1(\mathcal{L}(1,0)-\mathcal{L}(1,1))}\}$, 这与似然比检验 (LRT) 相同: Likelihood $L(X) := \frac{p(X|Y=1)}{p(X|Y=0)}$, 形式为 $f(x) = 1\{L(x) \geq \eta\}$.
- Confusion table:

	$Y = 0$	$Y = 1$
$\hat{Y} = 0$	true negative	false negative
$\hat{Y} = 1$	false positive	true positive

Ture Positive Rate: $\text{TPR} = \mathbb{P}(\hat{Y} = 1|Y = 1)$; False Negative Rate: $\text{FNR} = 1 - \text{TPR}$, type II error; False Positive Rate: $\text{FPR} = \mathbb{P}(\hat{Y} = 1|Y = 0)$, type I error; True Negative Rate: $\text{TNR} = 1 - \text{FPR}$. Precision: $\mathbb{P}(Y = 1|\hat{Y} = 1) = \frac{p_1 \text{TPR}}{p_0 \text{FPR} + p_1 \text{TPR}}$. F_1 -score: F_1 is the harmonic mean of precision and recall, which can be written as $F_1 = \frac{2\text{TPR}}{1 + \text{TPR} + \frac{p_0}{p_1} \text{FPR}}$.

- Optimization: maximize TPR subject to $\text{FPR} \leq \alpha, \alpha \in [0, 1]$. Randomized rule: Q return 1 with probability $Q(x)$ and 0 with probability $1 - Q(x)$. Maximize $\mathbb{E}[Q(x)|Y = 1]$ subject to $\mathbb{E}[Q(x)|Y = 0] \leq \alpha$. Suppose the likelihood functions $p(x|y)$ are continuous. Then the optimal predictor is a deterministic LRT (N-P lemma). (Proof: Let η be the threshold for an LRT such that the predictor $Q_\eta(x) = 1\{\alpha(x) \geq \eta\}$ has $\text{FPR} = \alpha$. Such an LRT exists because likelihood are continuous. Let β denote the TPR of Q_η . Prove that Q_η is optimal for risk minimization problem corresponding to the loss functions $\mathcal{L}(0, 1) = \eta \frac{p_1}{p_0}, \mathcal{L}(1, 0) = 1, \mathcal{L}(1, 1) = \mathcal{L}(0, 0) = 0$ since $\frac{p_0(\mathcal{L}(0,1)-\mathcal{L}(0,0))}{p_1(\mathcal{L}(1,0)-\mathcal{L}(1,1))} = \frac{p_0 \mathcal{L}(0,1)}{p_1 \mathcal{L}(1,0)} = \eta$. Under these loss functions, the risk of Bayes predictor for Q is $\mathcal{R}_Q = p_0 \text{FPR}(Q) \mathcal{L}(0, 1) + p_1(1 - \text{TPR}(Q)) \mathcal{L}(1, 0) = p_1 \eta \text{FPR}(Q) + p_1(1 - \text{TPR}(Q))$. Now let Q be any other rule with $\text{FPR}(Q) \leq \alpha, \mathcal{R}_{Q_\eta} = p_1 \eta \alpha + p_1(1 - \beta) \leq p_1 \eta \text{FPR}(Q) + p_1(1 - \text{TPR}(Q)) \leq p_1 \eta \alpha + p_1(1 - \text{TPR}(Q)) \Rightarrow \text{TPR}(Q) \leq \beta$)
- ROC (Receiver operating character) curve: y -axis is TPR and x -axis is FPR. Proposition: (1) The points $(0, 0)$ and $(1, 1)$ are on the ROC curve; (2) The ROC must lie above the main diagonal; (3) The ROC curve is concave. (Proof: (2): Fix $\alpha \in (0, 1)$ and consider a randomized rate $\text{TPR} = \text{FPR} = \alpha, Q(x) \equiv \alpha$; (3): Consider two rules $(\text{FPR}(\eta_1), \text{TPR}(\eta_1))$ and $(\text{FPR}(\eta_2), \text{TPR}(\eta_2))$. If we flip a biased coin and use the first rule with probability t and use the second rule with probability $1 - t$. Then this yields a randomized rule with $(\text{FPR}, \text{TPR}) = (t\text{FPR}(\eta_1) + (1 - t)\text{FPR}(\eta_2), t\text{TPR}(\eta_1) + (1 - t)\text{TPR}(\eta_2))$. Fixing $\text{FPR} \leq t\text{FPR}(\eta_1) + (1 - t)\text{FPR}(\eta_2)$, $\text{TPR} \geq t\text{TPR}(\eta_1) + (1 - t)\text{TPR}(\eta_2)$.)

3 马尔可夫决策过程

- Markov Decision Processes (MDPs): Five elements: decision epoches, states, actions, transition probabilities and rewards. (1) Decision epoches: Let T denote the set of decision epoches, discrete: $\{1, 2, \dots, N\}$; continuous: $[0, N]$; $N < / = \infty$: finite or infinite. (2) State and action sets: decision epoch $t \in T$, the system occupies a state $S_t \in \mathcal{S}$, the decision maker $a \in \mathcal{A}$. (3) Reward and transition probabilities: t , in state s , choose action a , (i) the decision maker receives a reward $r_t(s, a)$, (ii) the system state at the next decision epoch is determined by the probability distribution $p_t(\cdot|s_t, a)$.
- Decision rules: Prescribe a procedure for action selection in each state at a specified decision epoch. Four cases: (1) Markovian and Deterministic: $\delta_t : \mathcal{S} \rightarrow \mathcal{A}$; (2) M and Randomized: $\delta_t : \mathcal{S} \rightarrow \Delta(\mathcal{A})(q_{\delta_t(s)}(a))$; (3) History-dependent and D: $h_t = (s_1, a_1, \dots, s_{t-1}, a_{t-1}, s_t) = (h_{t-1}, a_{t-1}, s_t), \mathcal{H}_1 = \mathcal{S}, \mathcal{H}_2 = \mathcal{S} \times \mathcal{A} \times \mathcal{S}, \dots, \delta_t : \mathcal{H}_t \rightarrow \mathcal{A}$; (4) HR: $\delta_t : \mathcal{H}_t \times \Delta(\mathcal{A})$. A policy $\pi = (\delta_1, \delta_2, \dots, \delta_{N-1})$ is stationary if $\delta_1 = \delta_2 = \dots = \delta$ for $t \in T$.
- Let $\pi = (\delta_1, \dots, \delta_{N-1})$ in HR and $R_t := r_t(X_t, Y_t)$ denote the random reward, $R_N := r_N(X_N)$, $R := (R_1, \dots, R_N)$. The expected total reward $U_N^\pi(s) := \mathbb{E}^\pi\{\sum_{t=1}^{N-1} r_t(X_t, Y_t) + r_N(X_N) | X_1 = s\}$. Assume $|r_t(s, a)| \leq M < \infty$ for all $(s, a) \in \mathcal{S} \times \mathcal{A}$. Optimal policy: $U_N^*(s) \geq U_N^\pi(s), s \in \mathcal{S}$. ε -optimal policy: $U_N^{\pi^*}(s) + \varepsilon > U_N^\pi(s), s \in \mathcal{S}$. The value of the MDP: $U_N^*(s) = \sup_{\pi \in \mathcal{D}^{\text{HR}}} U_N^\pi(s), s \in \mathcal{S}$.

- Finite-Horizon Policy Evaluation: $V_t^\pi(h_t) = \mathbb{E}^\pi \left\{ \sum_{k=t}^{N-1} r_k(X_k, Y_k) + r_N(X_N) | h_t \right\}$, $V_N^\pi(h_N) = r_N(s)$, $\pi \in \mathcal{D}^{\text{HD}}$. 由重期望公式, $V_t^\pi(h_t) = r_t(s_t, \delta_t(h_t)) + \mathbb{E}_{h_t}^\pi V_{t+1}^\pi(h_t, \delta_t(h_t), X_{t+1}) = r_t(s_t, \delta_t(h_t)) + \sum_{j \in \mathcal{S}} V_{t+1}^\pi(h_t, \delta_t(h_t), j) \mathbb{P}(j | s_t, \delta_t(h_t))$. Consider randomness (i.e. $\pi \in \mathcal{D}^{\text{HR}}$): $V_t^\pi(h_t) = \sum_{a \in \mathcal{A}} q_{\delta_t(h_t)}(a) \{r_t(s_t, a) + \sum_{j \in \mathcal{S}} V_{t+1}^\pi(h_t, a, j) \mathbb{P}(j | s_t, a)\}$. Computational complexity: let $K = |\mathcal{S}|$, $L = |\mathcal{A}|$, at decision epoch t , $K^{t+1}L^t$ histories, $K^2 \sum_{i=0}^{N-1} (KL)^i$ multiplications. If $\pi \in \mathcal{D}^{\text{MD}}$, $V_t^\pi(s_t) = r_t(s_t, \delta_t(s_t)) + \sum_{j \in \mathcal{S}} V_{t+1}^\pi(j) \mathbb{P}(j | s_t, \delta_t(s_t))$, only $(N-1)K^2$ multiplications. On the other hand, given π , this yields a valid and accurate calculation method for $U_N^\pi(s)$.
- The Bellman Equations: Let $V_t^*(h_t) = \sup_{\pi \in \mathcal{D}^{\text{HR}}} V_t^\pi(h_t)$. The optimality equations: $V_t(h_t) = \sup_{a \in \mathcal{A}} \{r_t(s_t, a) + \sum_{j \in \mathcal{S}} V_{t+1}(h_t, a, j) \mathbb{P}(j | s_t, a)\}$ for $t = 1, 2, \dots, N-1$ and $h_t = (h_{t-1}, a_{t-1}, s_t) \in \mathcal{H}_t$. For $t = N$, $V_N(h_N) = r_N(s_N)$. Suppose V_t is a solution and V_N satisfies $V_N(h_N) = r_N(s_N)$. Then $V_t(h_t) = V_t^*(h_t)$ for all $h_t \in \mathcal{H}_t$, $t = 1, \dots, N$ and $V_1(s_1) = V_1^*(s_1) = U_N^*(s_1)$ for all $s_1 \in \mathcal{S}$. (Proof: Two parts. First prove $V_n(h_n) \geq V_n^*(h_n)$ for all $h_n \in \mathcal{H}_n$. By induction: $N : V_N(h_N) = r_N(s_N) = V_N^*(h_N)$ for all h_N, π . Now assume that $V_t(h_t) \geq V_t^*(h_t)$ for all $h_t \in \mathcal{H}_t$ for $t = n+1, \dots, N$. Let $\pi' = (\delta'_1, \dots, \delta'_{N-1})$ be an arbitrary policy in \mathcal{D}^{HR} . For $t = n$, the Bellman equations $V_n(h_n) = \sup_{a \in \mathcal{A}} \{r_n(s_n, a_t) + \sum_{j \in \mathcal{S}} \mathbb{P}(j | s_n, a) V_{n+1}(h_n, a, j)\} \geq \sup_{a \in \mathcal{A}} \{r_n(s_n, a) + \sum_{j \in \mathcal{S}} \mathbb{P}_n(j | s_n, a) V_{n+1}^*(h_n, a, j)\} \geq \sup_{a \in \mathcal{A}} \{r_n(s_n, a) + \sum_{j \in \mathcal{S}} \mathbb{P}_n(j | s_n, a) V_{n+1}^{\pi'}(h_n, a, j)\} \geq V_n^{\pi'}(h_n)$. Second prove for any $\varepsilon > 0$, there exists a $\pi \in \mathcal{D}^{\text{HD}}$ for which $V_n^{\pi'}(h_n) + (N-n)\varepsilon \geq V_n(h_n) \Rightarrow V_n^*(h_n) + (N-n)\varepsilon \geq V_n^{\pi'}(h_n) + (N-n)\varepsilon \geq V_n(h_n) \geq V_n^*(h_n)$. Construct a policy $\pi' = (\delta'_1, \dots, \delta'_{N-1})$ by choosing $\delta'_n(h_n)$ to satisfy $r_n(s_n, \delta'_n(h_n)) + \sum_{j \in \mathcal{S}} \mathbb{P}_n(j | s_n, \delta'_n(h_n)) V_{n+1}(h_n, \delta'_n(h_n), j) + \varepsilon \geq V_n(h_n)$. By induction: $N : V_N^{\pi'}(h_N) = V_N(h_N)$. Assume that $V_t^{\pi'}(h_t) + (N-t)\varepsilon \geq V_t(h_t)$ for $t = n+1, \dots, N$. For $t = n$, $V_n^{\pi'}(h_n) = r_n(s_n, \pi'_n(h_n)) + \sum_{j \in \mathcal{S}} \mathbb{P}_n(j | s_n, \delta_n^{\pi'}(h_n)) V_{n+1}^{\pi'}(h_n, \delta_n^{\pi'}(h_n), j) \geq V_n(h_n) - (N-n)\varepsilon$. The equations yield that $\delta_t^*(h_t) \in \arg \max_{a \in \mathcal{A}} \{r_t(s_t, a) + \sum_{j \in \mathcal{S}} \mathbb{P}_t(s_t, a) V_{t+1}^*(h_t, a, j)\}$, which means it is HD, i.e. $U_N^*(s) = \sup_{\pi \in \mathcal{D}^{\text{HR}}} U_N^\pi(s) = \sup_{\pi \in \mathcal{D}^{\text{HD}}} U_N^\pi(s) \stackrel{?}{=} \sup_{\pi \in \mathcal{D}^{\text{MD}}} U_N^\pi(s)$.
- Let $V_t^*, t = 1, \dots, N$ be solutions of Bellman Equations. Then (a) For each $t = 1, \dots, N$, $V_t^*(h_t)$ depends on h_t only through s_t ; (b) For any $\varepsilon > 0$, there exists an ε -optimal policy which is D and M; (c) Max can be achieved, it is optimal, which is MD. (Proof: (a): By induction, $V_N^*(h_N) = V_N^*(h_{N-1}, a_{N-1}, s) = r_N(s)$ for all $h_{N-1} \in \mathcal{H}_{N-1}$. Assume (a) is valid for $t = n+1, \dots, N$. Then $V_n^*(h_n) = \sup_{a \in \mathcal{A}} \{r_t(s_t, a) + \sum_{j \in \mathcal{S}} \mathbb{P}_t(j | s_t, a) V_{t+1}^*(j)\} = V_n^*(s_t)$.
- Backward Induction (Dynamic Programming) Algorithm: 1. Set $t = N$ and $V_N^*(s_N) = r_N(s_N)$ for all $s_N \in \mathcal{S}$; 2. Substitute $t-1$ for t and compute $V_t^*(s_t)$ for each $s_t \in \mathcal{S}$: $V_t^*(s_t) = \max_{a \in \mathcal{A}} \{r_t(s_t, a) + \sum_{j \in \mathcal{S}} \mathbb{P}_t(j | s_t, a) V_{t+1}^*(s_t)\}$, set $\mathcal{A}_{s_t} = \arg \max_{a \in \mathcal{A}} \{r_t(s_t, a) + \sum_{j \in \mathcal{S}} \mathbb{P}_t(j | s_t, a) V_{t+1}^*(s_t)\}$; 3. If $t = 1$, stop. Otherwise return to Step 2.
- Other remarks: (1) At time t , specialized \mathcal{S}_t and \mathcal{A}_s , special structure for r_t and \mathbb{P}_t ; (2) $K = |\mathcal{S}|$ and $L = |\mathcal{A}|$, at each t , only $(N-1)LK^2$ multiplications, ease computation and storage cost (because there are $(L^K)^{N-1}$ DM policies).