Azuma's Ineq.) Let $\{S_k, k=0,1,2,\cdots\}$ be a martingale. w/ $|S_k - S_{k-1}| \le c_k$ a.s.

$\mathbb{E}[S_{k+1} | S_0, \cdots, S_k] = S_k$

Then. fix any N. w.p. $\ge 1-\delta$.

$$|S_N - S_0| \le \sqrt{2\left(\sum_{k=1}^{N} c_k^2\right)\log\frac{2}{\delta}}.$$

Indep: $X_1, X_2, \cdots, X_n$, $\left|\frac{\sum_{k=1}^n X_k}{n} - \frac{\sum_{k=1}^n \mathbb{E}[X_k]}{n}\right|$

$$= \frac{1}{n}\left|\sum_{k=1}^{n}\underbrace{(X_k - \mathbb{E}[X_k])}_{0 \text{ mean}}\right| \to O(\sqrt{n}).$$

$S_0 = 0$, $S_n = \sum_{k=1}^{n}(X_k - \mathbb{E}[X_k])$.

$$\mathbb{E}[S_{n+1} | S_0, \cdots, S_n] = \mathbb{E}[S_{n+1} - S_n + S_n | S_{0:n}]$$

$$= \mathbb{E}[X_{n+1} - \mathbb{E}[X_{n+1}] | S_{0:n}] + S_n$$

$$= S_n.$$

Example: MAB w arms a, a'.

reward dist: R, R' $[0,1]$.

mean : μ, μ'.

Alg: toss a coin $\begin{cases} \text{always pulls } a \\ \text{always pulls } a'. \end{cases}$

$a_1, \; a_2, \; a_3 \cdots \cdots a_N$

$r_1, \; r_2, \; r_3, \cdots \cdots, r_N.$

① $\sum_{i=1}^{N} (r_i - \mu_{a_i})$

② $\sum_{i=1}^{N} (r_i - \underbrace{\mathbb{E}[r_i]}_{\frac{\mu+\mu'}{2}})$.

General:

$Z_1 \quad Z_2 \quad Z_3 \cdots \quad Z_n$

$\underline{\underline{Y_1}} \quad \underline{\underline{Y_2}} \quad \underline{\underline{Y_3}} \cdots \quad \underline{\underline{Y_n}}.$

$$\sum_{i=1}^{n} (\underline{\underline{Y_i}} - \underline{\mathbb{E}[Y_i \mid Z_{1:i}, Y_{1:i-1}]}),$$

$\Delta \qquad \|$

MAB:

$\pi_3 \in \Delta(A)$. $\mathbb{E}[Y_i \mid \text{Bottleneck}, \varepsilon_{1:i}, Y_{1:i-1}]$

$a_1 \quad a_2 \quad a_3, \quad \dots \quad a_N$

$r_1 \quad r_2 \quad r_3 \quad\quad\quad r_N.$

(1) $\sum_{i=1}^{N} (r_i - \mu_{a_i})$. ← bottleneck: a_i.

$a_{1:i}, r_{1:i} \perp r_{i+1} \mid a_{i+1}$

(2) $\sum_{i=1}^{N} \left(r_i - \sum_a \pi_i(a) \cdot \mu_a \right)$,

$\mathbb{E}[r_i \mid r_1, \dots, r_{i-1}, \boxed{\quad}]$.

$\forall t$. w.p. $\geq 1-\delta$.

$\rightarrow \left\| \sum_{i=1}^{t-1} x_i \varepsilon_i \right\|_{\Lambda_t^{-1}} \leq O\left(V_{max}\sqrt{d + \log\frac{1}{\delta}}\right)$.

$x_{1:i}, \varepsilon_{1:i-1}$,

$\mathbb{E}[x_i \varepsilon_i \mid x_i]$

$= x_i \mathbb{E}[\varepsilon_i]$

$\mathbb{E}[x_i \varepsilon_i \mid x_{1:i}, \varepsilon_{1:i-1}] = 0$?

Ridge Regression

$t = 1, 2, 3, \dots, T.$

- Nature chooses $x_t \in \mathbb{R}^d$. $(\|x_t\| \le 1)$.
- Noisy Label: $\underbrace{y_t}_{[0, V_{max}]} = x_t^T \boxed{\theta^*} + \underline{\varepsilon_t} \to$ 0-mean noise.

Goal: at time t, use $x_{1:t-1}$, $y_{1:t-1}$ to predict $y(x) = x^T \theta^*$ $\forall x$.

Alg: $\hat{\theta}_t = \underset{\theta \in \mathbb{R}^d}{\operatorname{argmin}} \sum_{i=1}^{t-1} (x_i^T \theta - y_i)^2 + \|\theta\|^2.$

$\Leftrightarrow$ Define: $\Lambda_t := I + \sum_{i=1}^{t-1} x_i x_i^T \in \mathbb{R}^{d \times d}.$

$$\hat{\theta}_t = \Lambda_t^{-1} \sum_{i=1}^{t-1} (x_i y_i).$$

$$\Rightarrow \hat{y}_t(x) = x^T \hat{\theta}_t = x^T \Lambda_t^{-1} \left(\sum_{i=1}^{t-1} x_i y_i \right).$$

Fix any t, w.p. $\ge 1 - \delta$. $\forall x \in \mathbb{R}^d$

$$\rightarrow \left| \hat{y}_t(x) - \underset{x^T\theta^*}{y(x)} \right| \le \boxed{\|x\|_{\Lambda_t^{-1}}} \, O\left(V_{max} \sqrt{d + \log \frac{1}{\delta}}\right)$$

$\tilde{x} = a x + b x'.$

$x \qquad x' \quad \Rightarrow \quad y(\tilde{x}) = a\, y(x) + b\, y(x').$

$\|x\|_A = \sqrt{x^T A x}$. for PSD A.

$\|x\|_{\Lambda_t^{-1}} = \sqrt{x^T \Lambda_t^{-1} x}$.

Online prediction: $t = 1, 2, \cdots, T$.

- Nature x_t.
- Learner predicts $\hat{y}_t = \hat{y}_t(x_t)$
- $y_t = x_t^T \theta^* + \varepsilon_t$ is revealed.

$$\text{Regret}_T := \sum_{t=1}^{T} |\hat{y}_t - y(x_t)| = \tilde{O}(\sqrt{T})$$

---

w.p. $\geq 1-\delta$, for all $t \in [1, T]$

$$|\hat{y}_t(x_t) - y(x_t)| \leq \|x_t\|_{\Lambda_t^{-1}} \cdot \tilde{O}\left(V_{max}\sqrt{d + \log\frac{1}{\delta}}\right)$$

hide $\frac{d, V_{max}}{\log T}$.

$$\sum_t |\hat{y}_t(x_t) - y(x_t)| \leq \tilde{O}(\cdot) \cdot \sum_{t=1}^{T} \|x_t\|_{\Lambda_t^{-1}}$$

$$\sum_{t=1}^{T} \sqrt{x_t^T \Lambda_t^{-1} x_t} \le \sqrt{T} \cdot \sqrt{\sum_{t=1}^{T} x_t^T \Lambda_t^{-1} x_t}.$$

Lemma (Elliptical Potential Lemma).

$$\sum_{t=1}^{T} x_t^T \Lambda_t^{-1} x_t \le 2\log \det(\Lambda_{T+1}) \le 2d\log(T+1).$$

① $\|x_t\| \le 1$.

② $\Lambda_t = I + \sum_{i=1}^{t-1} x_i x_i^T$.

$\min_{\|u\|=1} u^T A u$.

$= 1 + \sum_{i=1}^{t-1} u^T x_i x_i^T u$

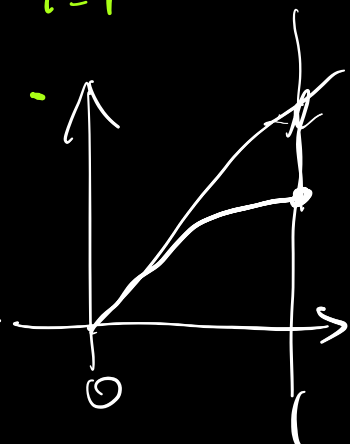
$= 1 + \sum_{i=1}^{t-1} (u^T x_i)^2$

≥ 1.

Proof: $\forall t$. $x_t^T \Lambda_t^{-1} x_t$

$\le x_t^T I x_t \le 1$.

$\forall z \in [0,1]$. $z \le 2\log(1+z)$.



$$\det(\Lambda_{t+1}) = \det(\Lambda_t + x_t x_t^T).$$

$$= \det\left(\Lambda_t^{1/2}\left(I + \Lambda_t^{-1/2} x_t x_t^T \Lambda_t^{-1/2}\right)\Lambda_t^{1/2}\right)$$

$$= \det(\Lambda_t^{1/2}) \det(\Lambda_t^{1/2}).$$

$$\det(I + \underbrace{\Lambda_t^{-1/2} x_t}_{A} \underbrace{x_t^T \Lambda_t^{-1/2}}_{A^T})$$

Λ real sym. PSD.

$\Lambda = U^T \Sigma U$.

$\Lambda^{1/2} = U^T \Sigma^{1/2} U$

$$X_t \Lambda_t^{-1} X_t^T = AA^T.$$

$$\Lambda^{-1} = U^T \Sigma^{-1} U.$$

$$\Lambda^{1/2} \Lambda^{1/2} = U^T \Sigma^{1/2} U U^T \Sigma^{1/2} U = \Lambda.$$

$$U^T = U^{-1}.$$

$$\det(I + A^T A) = \det(I + AA^T).$$

$$A^T A x = \lambda x.$$

$$(AA^T)(Ax) = \lambda (Ax). \quad AA^T \quad A^T A$$

$$= \det(\Lambda_t) \det(I + X_t^T \Lambda_t^{-1} X_t).$$

$$= \det(\Lambda_t) (1 + x^T \Lambda_t^{-1} x_t).$$

$$\log \frac{\det \Lambda_{t+1}}{\det \Lambda_t} = \log(1 + x^T \Lambda_t^{-1} x_t)$$

$$\geq \frac{1}{2} x_t^T \Lambda_t^{-1} x_t.$$

$$\Rightarrow \sum_{t=1}^{T} x_t^T \Lambda_t^{-1} x_t \leq 2 \cdot \sum_{t=1}^{T} \log \frac{\det \Lambda_{t+1}}{\det \Lambda_t}$$

$$x^T A x \leq 1.$$

$$= 2 \cdot \log \frac{\det \Lambda_{T+1}}{\det I}$$

$$= 2 \cdot \log \det \Lambda_{T+1}.$$

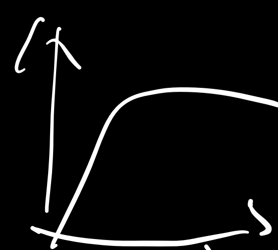
$$\det \Lambda_{T+1} \le \sigma_{max}(\Lambda_{T+1})^d \le (T+1)^d.$$

$$\sigma_{max}(\Lambda_{T+1}) = \max_{\|u\|=1} u^T \Lambda_{T+1} u$$

$$= \max_{\|u\|=1} u^T \sum_{t=1}^{T} (I + x_t x_t^T) u.$$

$$\le T+1.$$

$$\Lambda_t = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix}.$$



$$\sum x_t^T \Lambda_t^{-1} x_t = \sum_{i=1}^{d} \left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n_i} \right)$$

$$\approx \sum_{i=1}^{d} \log n_i.$$

$$\simeq d \cdot \log \frac{T}{d}$$

$\boxed{\text{Covering}}$ $\quad \mathcal{F} \quad \log|\mathcal{F}|$.

Let $f \in [0,1]$, $\forall f \in \mathcal{F}$.

$$\left| \mathbb{E}[f(x)] - \frac{1}{n}\sum_{i=1}^{n} f(x_i) \right| \le \; ?$$

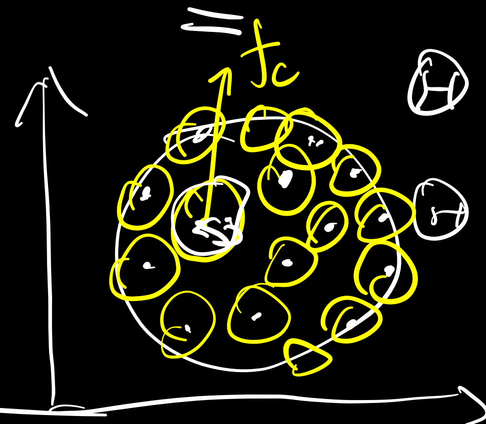
if $|\mathcal{F}| = \infty$?

Def [l_∞ covering #].

$\underline{\mathcal{F}_\varepsilon}$ is an ε-cover of $\mathcal{F}$ (under l_∞).

if: $\forall f \in \mathcal{F}$. $\exists f_c \in \mathcal{F}_\varepsilon$, $\|f - f_c\|_\infty \le \varepsilon$.

$N_\varepsilon := N_\varepsilon(\mathcal{F})$ is the size of smallest cover of $\mathcal{F}$.



$\|\theta\|_\infty \le C$.

$$N_\varepsilon(\mathcal{F}) \le \left(\frac{2C}{\varepsilon/\sqrt{d}} \right)^d.$$

$$\log N_\varepsilon(\mathcal{F}) \approx d \log(\;\;)$$

$\|\theta_c - \theta\|_\infty \le \varepsilon_0$.

$\|\theta_c - \theta\|_2 \le \varepsilon_0 \sqrt{d}$.

$$\|f - f_c\|_\infty = \max_{x \in \mathcal{X}} (x^T\theta - x^T\theta_c)$$

$$\le \|X\|_2 \cdot \|\theta - \theta_c\|_2$$

$$\le \quad 1 \cdot \|\theta - \theta_c\|_2$$

$$\le \quad \sqrt{d}\,\varepsilon_0 = \varepsilon.$$

w.p. $\geq 1-\delta$, $\forall f_c \in \mathcal{F}_\varepsilon$.

$$\left| \mathbb{E}[f_c] - \hat{\mathbb{E}}[f_c] \right| \le \sqrt{\frac{1}{n} \log \frac{2N_\varepsilon}{\delta}} = \frac{\varepsilon}{3}.$$

$\forall \; f \in \mathcal{F}$.

$$\left| \mathbb{E}[f] - \hat{\mathbb{E}}[f] \right|$$

$$\le \left| \mathbb{E}[f_c] - \hat{\mathbb{E}}[f_c] \right| \quad \frac{\varepsilon}{3}$$

$$+ \left| \mathbb{E}[f_c] - \mathbb{E}[f] \right| \quad \frac{\varepsilon}{3} \quad \le \|f - f_c\|_\infty.$$

$$+ \left| \hat{\mathbb{E}}[f_c] - \hat{\mathbb{E}}[f] \right|. \quad \frac{\varepsilon}{3} \quad \le \varepsilon.$$