NeuRIPS 2019

– DAY 1 –

Tutorial 1

E. Khan (AIP RIKEN)

– DL w. Bayesian Principles –

Tutorial 2

Gretton, Sutherland, Jitkrittum

– Interpretable Comparison of Distributions and Models –

Tutorial 3

K. Hoffmann (Microsoft)

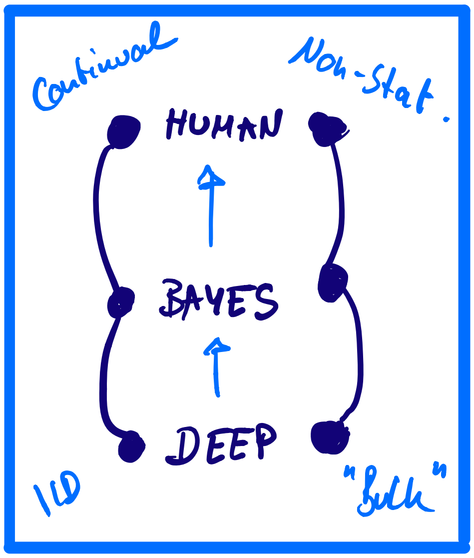
– RL: Past, Present, Future –

Keynote 1

C. Kidd (Berkeley)

– How to Know –

Emtiyaz Khan (RIKEN) – DL with Bayesian Principles –



BAYESIAN PRINCIPLE

↓

POSTERIOR APPROX.

↓

* SGD / Adam
* Exact, Laplace, VI

⇒ BRING TOGETHER!

	DL	BL
COMPLEX DATA / TRAIN	✓	✗
SCALABLE	✓	✗
UNCERT.	✗	✓
INCREM.	✗	✓

DEEP LEARNING

→ START: Emp. Risk Min / Max. LH

$$\min_{\theta} \ell(D;\theta)$$

PRE-COND.

→ SGD: $\theta \leftarrow \theta - \rho H^{-1} \nabla_\theta \ell$

↳ Point Estimate ⇒ LOCAL

BAYESIAN LEARNING

→ Prior: $\theta \sim p(\theta)$

→ Score: $p(D|\theta)$ (via LH)

→ $p(\theta|D) = \dfrac{p(\theta)\,p(D|\theta)}{\int p(D|\theta)\,p(\theta)\,d\theta}$

↳ Weight ⇒ GLOBAL

COMPLEXITIES OF POSTERIOR APPROX.

① SGD ($H = I$)

⇒ Gaussian w. fixed Covariance

② Newton

⇒ Multivariate Gaussian

③ Ensemble

⇒ Gaussian Mixture

Ⓘ CHOOSE POSTERIOR APPROX.

Exp. Family

$q(\theta) \propto \exp(\lambda^T T(\theta))$ w. $\mu = \mathbb{E}_q[\theta]$

Natural Param. Sufficient Stat. Exp. Param.

$$\min_{q \in Q} \mathbb{E}_{q(\theta)}[\ell(\theta)] - \mathcal{H}(q)$$

ENTROPY: PREVENT COLLAPSE

BAYES. L. RULE:

$\lambda \leftarrow \lambda - \rho \nabla_\mu [\mathbb{E}_q[\ell(\theta)] - \mathcal{H}(q)]$

Ⓘⓘ LET GLOBALITY GO AWAY

⇒ Khan et al. 18' – ICML

- RMSprop ⇒ Can be derived using Multivariate Gaussian w. diag Σ

BAYES = OPTIM: $p(\theta|D) = \arg\min_{q \in \mathcal{P}} \mathbb{E}_{q(\theta)}[\ell(\theta)] - \mathcal{H}(q)$

$\mathcal{P}$: all; $q(\theta)$: approx.; $\ell(\theta)$: $-\log p(D|\theta) \times p(\theta)$; $\mathcal{H}(q)$: entropy

↳ View HMM, Kalman Filter, etc. from exact same perspective

⇒ Gradient update given correct approximation! → PRINCIPLE → NOT VI ONLY!

I) BUILDING BETTER DL MODELS: UNCERTAINTY ESTIMATION

A] SGD-Based

- MC Dropout – Gal & Gahrani 16'
- SWAG – Maddox et al. 19'
- Laplace – Ritter et al. 18'

⇒ PROBLEM: NOT FLEXIBLE!

B] VI Methods

- Flexible q distr.
- Graves 11'

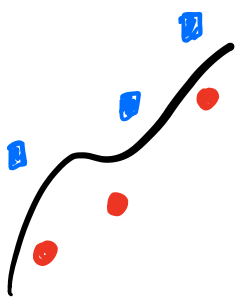
⇒ PROBLEM: HARD TO SCALE!

↳ VOGN: Osawa et al. 19' (NeurIPS) ⇒ Variational Online Gauss-Newton

* Adam-like optimizer: Sampling around solution space to estimate uncertainty

↳ Challenge: Non-convexity → local approx. ⇒ Only local uncertainty

II) BUILDING BETTER DL MODELS: DATA IMPORTANCE



WEIGHT SPACE ← DNN → Global

↓ POSTERIOR APPROX. ↓

FCT. SPACE ← GP → Local

⇒ DUAL PDV via linear basis functions

↳ Importance via sorting of variances of datapoints

III) BUILDING BETTER DL MODELS: LIFELONG LEARNING (→ Active DL)

Continual learning → Regularization-Based (EWC, SI, VCL)

- Classification: Past classes not visited

↳ BROKEN PRINCIPLE: BETTER APPROX. ≠ BETTER RESULTS

↳ VOGN can fix problem!

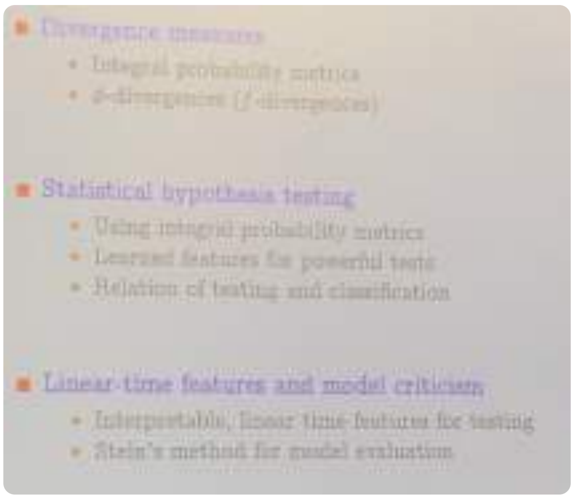
Jitkrittum (II), Sutherland (II) & Gretton (I)

– Interpretable Comparison of Distributions & Models –

Scenario \ Distr.	P	Q
(I)	Samples	Samples
(II)	Model	Samples

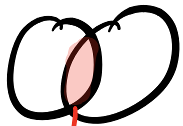
⇒ Example: GANs

OUTLINE



(I) DIVERGENCE MEASURES

(A) P – Q
↓
Difference
↓
Integral Prob. Metrics



Total variation

(B) P / Q
↓
Ratio
↓
ϕ-divergences

(A) I Max. Mean Discrepancies (MMD)

$$\sup_{\|f\| \leq 1} \mathbb{E}_P f(X) - \mathbb{E}_Q f(Y)$$

↳ RKHS: Smoothness / Fct. lives in features

↳ Different witness functions ⇒ Metrics

↳ MMD: Infinite many features ⇒ Kernel Trick

$$f^* \propto \mu_P - \mu_Q$$

(A) II Wasserstein-1 Distance

$$\sup_{\|f\|_L \leq 1} \mathbb{E}_P f(X) - \mathbb{E}_Q f(Y)$$

→ Quality of teacher signal depends on choice of kernel!

⇒ COMPARISON: What to use? WHE care!

↳ IPM more stable in terms of hyperparameter robustness

(B) Reverse KL

$$D_\phi(P,Q) = \int \phi\left(\frac{dP}{dQ}\right) dQ$$

→ In practice: lower Bound max.

≈ "KALE" divergence → different behaviour

↳ smoother feedback! ⇒ Can help with mode collapse

⇒ RIGHT SMOOTHING OF RATIO IS REQUIRED

(II) STATISTICAL HYPOTHESIS TESTING (Two Sample)

→ Example Problem: CIFAR-10 (P) vs. CIFAR 10.1 (Q) ⇒ Are P & Q equal?

→ Practice: Compute kernel matrix and sample average of $\widehat{MMD}^2$

↳ $P \neq Q$: Asymptotic normal
↳ $P = Q$: χ^2 distribution mixture
→ depends on kernel

] Hypothesis Test Construction
Choose c_α → false rejection rate

⇒ How to set c_α? → Permutation Testing ⇒ Mix Data Randomly!

⇒ How to choose kernel? → Exp. Quadratic ⇒ "Characteristic" for all σ
↳ power → 1 as $n \to \infty$

Only asymptotic ⇒ Finite case: σ matters

↳ Choose a meaningful representation! $\Phi: \mathcal{X} \to \mathbb{R}^d$

Φ: Prediction from Classifier // Φ: Hidden Layer Features (better generalization)

→ "Best Test": Data Adaptation of Kernel ⇒ Goal: Highest Power!

↳ $\max\limits_{k} \dfrac{MMD^2(P,Q)}{\sigma_{H_1}(P,Q)}$ ⇒ Trade-Off large MMD with confidence!
↳ Differentiable! ⇒ Autodiff Custom Loss Def.

learn kernel! → "Deep Kernel"

↳ Kernel as Classifier: $k_f(x,y) = 1/4 \ \mathbb{1}(f(x)>0) \ \mathbb{1}(f(y)>0)$

⇔ $MMD(P,Q) = |\text{accuracy} - 1/2|$

→ Interpretability ⇒ learned kernel + Large Witness Fct. Values

III LINEAR-TIME FEATURES & MODEL CRITICISM

→ Test location $V: \{v_j\}_{j=1}^J$ optimisation for witness function → Unnormalized Mean Embeddings (UME)

↳ Problem: runs in quadratic time

↳ Normalized Mean Embeddings (NME)

→ Optimisation ⇒ Where does GAN model do well?! → witness fct.

→ Explicit model P with Stein Discrepancy ⇒ 'Goodness-of-fit' Testing

density but no samples!

↳ $\sup\limits_{\|f\|_{\mathcal{F}} \leq 1} \left[\mathbb{E}_q T_p f - \mathbb{E}_p T_p f \right]$ → Stein Operator:

$$[T_p f](x) = \frac{1}{p(x)} \frac{d}{dx} \left(f(x)p(x)\right)$$

⇒ Kernel Stein Discrepancy (KSD)
↳ derive Stein Witness Fct.

Katja Hofmann (MSR) - Reinforcement Learning: Past, Present & Future Perspectives

$\boxed{\text{RL = SEQUENTIAL DECISION MAKING UNDER UNCERTAINTY}}$

ORIGINS - MARKOV DECISION PROCESS
→ Bellman (1953, 1954)

BEHAVIOUR POLICY
$\pi(a|s)$

$T(s_{t+1}|s_t, a_t)$
DYNAMICS

$R(r_{t+1}|s_t, a_t)$
REWARDS

⇒ Many decision choices
↳ state/action representation
⇒ Extensions: POMDPs / SMDPs

OPTIMALITY

Finite Horizon	Infinite Horizon	Average Reward
$\mathbb{E}\left[\sum_{t=0}^{h} r_t\right]$	$\mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r_t\right]$	$\lim_{h} \mathbb{E}\left(\frac{1}{h}\sum_{t=0}^{h} r_t\right)$

PERFORMANCE

Asympt. Convergence	PAC	Regret
$\pi_h \to \pi^*$ as $h \to \infty$	$P(N_{err} > F(.)) \leq \delta$	$\max_j \sum_t r_{t_j^*} - r_t < B$

DYNAMIC PROGRAMMING → Break problem into components
⇒ Based on Markov Assumption → BOOTSTRAP BACKUP!
↳ VALUE FCT. ESTIMATES

$$\boxed{\Rightarrow V^{\pi}(s_t) = R(s, \pi(s_t)) + \gamma \sum_{s_{t+1}} T(s_{t+1}|s_t, \pi(s_t))\, V^{\pi}(s_{t+1})}$$

δ: TD Error

↳ TEMPORAL DIFFERENCE (TD) LEARNING: $V(s_t) \leftarrow V(s_t) + \alpha\left[r_t + \gamma V(s_{t+1}) - V(s_t)\right]$

↳ Q-LEARNING: $Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha\left[r_t + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)\right]$

↳ FUNCTION APPROX.: $Q(s_t, a_t, \theta_k)$ ⇒ Instable: DEADLY TRIAD
- Off-Policy
- Bootstrapping
- Fct. Approx.

⇒ BREAKTHROUGH: DQNs (M

1. Experience Replay Buffer
2. Separate Target Networks
3. Clip TD error to [-1, 1] range

$$\boxed{\begin{aligned} \min_{\theta} \mathcal{L}_{MSBE} &= \mathbb{E}\left[Q(s, a; \theta_k) - y_k\right] \\ y_k &= r + \gamma \max_{a'} Q(s', a'; \theta^-) \end{aligned}}$$

EXTENSIONS: DOUBLE, APE-X, DUELING, AVERAGE, HER

EXPLORATION: OPTIMISTIC INIT, POSTERIOR SAMPLING, ε-Greedy, UCB, Softmax

⊙ Posterior Sampling for DQN ⇒ Bootstrapped DQN (Osband et al., 16')
↳ Uncertainty estimate via ensemble ⇒ Several Output Heads
↳ Osband et al., 18' ⇒ Systematic exploration via randomized prior fct.
↳ Janz et al., 19' ⇒ Successor Uncertainties based on Successor Features

---

REINFORCE: Policy Gradient Algo π_θ

* $R_i = \sum_{u=i}^{t} \gamma^{t-u} r_u$ ⇒ MC Returns
* $\hat{A}_i = R_i - b$ ⇒ Baseline Correction
* $\theta = \theta + \alpha \nabla_\theta \log \pi_\theta(a|s_i) \hat{A}_i$ ⇒ PG Update

⇒ Actor-Critic Methods: Learn value fct. to provide sample-efficient signals

↓

LEARNING SPEED vs. STABILITY

CPI	SURROGATE + MONOTONIC IMPROVEMENT
TRPO	APPROX. CPI w. TRUST REGION
PPO	KL PENALTY + CLIPPING
SAC	MAX ENTROPY RL

SAMPLE EFFICIENT RL

GENERALIZATION: Need for RL-specific regularizers
* Igl 19': Selective noise injection ⊕ info bottleneck

STRUCTURE: Meta-Learning ⇒ Augmented MDP
* Relationship via low-dim. embedding
* MAML: Inner + Outer loop ⇒ learn Init θ
* CAVIA: Separate context ⇒ Rapid adaptation

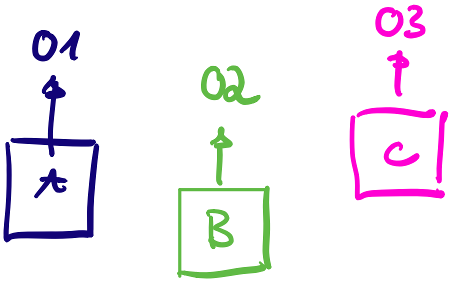
MODEL-BASED:
* PILCO - Deisenroth, Rasmussen 11'
* World Models - Ha, Schmidhuber 18'
* Bayesian NNs - Chua et al. 18'
* Meta-L. - Saemundson et al. 18', 19'

CHALLENGES:
* MARLO ⇒ 2 player mini-games
* Mine RL ⇒ Human priors

Celeste Kidd (UC Berkeley) – How to Know

DEVELOPMENTAL PERSPECTIVE → BELIEF FORMATION ⇒ ⊕ Behaviour Model

1 HUMANS CONTINUOUSLY FORM BELIEFS (Prob. Expectations)



→ Dirichlet - Multinomial ⇒ Surprise regressors ⇒ LOOK AWAY PROBS

↳ Goldilocks effect : ⇈ / ⇊ Surprise ⇒ Look aways

↳ Predictable ⟷ Surprising

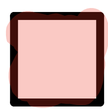
⇒ FIND / HIGHLIGHT DATA "AT THE EDGE" → Attention ⇒ Learning Progress

2 CERTAINTY DIMINISHES INTEREST

⇒ INTEREST DETERMINED BY WHAT YOU THINK YOU KNOW ⇒ NOT REALITY

3 CERTAINTY IS DRIVEN BY FEEDBACK

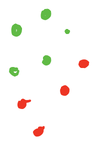
→ Open Mind : "Daxxyness" ⇒ META-COGNITION



↳ "Early stages of "making up your mind" is crucial to formation

4 LESS FEEDBACK MAY ENCOURAGE OVERCONFIDENCE

Nixxon → Concepts of politicians across population



⇒ t-SNE plot from Chinese Restaurant Process

→ Increased deviant concept awareness

5 HUMANS FORM BELIEFS QUICKLY

→ Importance of sequence of observation presentation

MAJOR ETHICAL CHALLENGES

→ NO NEUTRAL PLATFORMS

→ CHILDREN ARE SUSCEPTIBLE

→ GENDER DISCRIMINATION