Probabilistic Artificial Intelligence summary

Michael Etienne Van Huffel, Andrea Ghirlanda, Dr. Linda Maria de Cave

Probabilistic Artificial Intelligence summary created by *michavan@student.ethz.ch*, *aghirlanda@student.ethz.ch* and *ldecave@student.ethz.ch*

This summary has been written based on the Lecture 5263-5210-00 L Probabilistic Artificial Intelligence by Prof. A. Krause (Fall 2023). There is no guarantee for completeness and/or correctness regarding the content of this summary. Use it at your own discretion. The language used in this cheat sheet contains many abbreviations, which were necessary to adhere to the two-page limit imposed by the lecturer.

Basics (iid ⇔ uncorrelated, $S=-\log u$)
$P(X_{1:n}) = P(X_1)P(X_2|X_1)\cdots P(X_n|X_{1:n-1})$
Bayes: $P(X|Y)=\frac{P(X,Y)}{P(Y)}=\frac{P(Y|X)P(X)}{P(Y)}$
$\mathrm{var}(X\pm Y)=\mathrm{var}(X)+\mathrm{var}(Y)\pm 2\mathrm{cov}(X,Y)$
$\mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]=\mathbb{E}[XY]-\mathbb{E}[X]\mathbb{E}[Y]$
Gauss: $\frac{1}{\sqrt{(2\pi)^d|\Sigma|}}\exp(-\frac12(x-\mu)^T\Sigma^{-1}(x-\mu))$
Cond: $P(X_A|X_B=x_B)=\mathcal{N}(\mu_{A|B},\Sigma_{A|B})$
$\mu_A+\Sigma_{AB}\Sigma_{BB}^{-1}(x_B-\mu_B)$, $\Sigma_{AA}-\Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA}$
$H(\prod p_i)=\sum_i H(p_i)$; $H(N(\mu,\Sigma))=\frac12 ln|2\pi e\Sigma|$
$H(p,q)=H(p)+H(q|p)$, $H[p||q]=H[p]+KL(p||q)$
Convex: g(x) $\smile\Leftrightarrow x_1,x_2\in\mathbb{R},\lambda\in[0,1]$: $g''(x)>0$; $g(\lambda x_1+(1-\lambda)x_2)\le\lambda g(x_1)+(1-\lambda)g(x_2)$ **Egg:** $g\smile$: $g(E[X])\le E[g(X)]$
Bayes: *Prior:* $p(\theta)$, *Like:* $p(y_{1:n}|x_{1:n},\theta)=\prod_{i=1}^n p(y_i|x_i,\theta)$, *Post:* $p(\theta|x_{1:n},y_{1:n})=\frac1Z p(\theta)\prod_{i=1}^n p(y_i|x_i,\theta)$, $Z=\int(*)d\theta$, *Pred:* $p(y^*|x^*,x_{1:n},y_{1:n})=\int p(y^*|x^*,\theta)p(\theta|x_{1:n},y_{1:n})d\theta$

BLR (Gauss prior/noise, ass. same as Ridge)
BayI $w_{ls}=w_{MLE}, w_{ridg}=w_{MAP}(\lambda=\sigma_n^2/\sigma_p^2)$
Test x^*, $f^*=w^Tx^*$, $y^*=f^*+\epsilon$, $\epsilon\sim\mathcal N(0,\sigma_n^2)$
$p(\mathbf w)=\mathcal N(0,\sigma_p^2\mathbf I)$, $p(\mathbf w|X,y)=\mathcal N(\mathbf w;\overline\mu,\overline\Sigma)$
$p(y_i|x_i,w,\sigma_n)=\mathcal N(y_i;w^Tx_i,\sigma_n^2)$, $\overline\mu=\sigma_n^{-2}\overline\Sigma X^Ty$
$\overline\Sigma=(\sigma_n^{-2}\mathbf X^TX+\sigma_p^{-2}I)^{-1}$, $p(f^*y^*|\mathbf X,\mathbf y,\mathbf x^*)=\mathcal N(\mathbf x^{*T}\overline\mu,\mathbf x^{*T}\overline\Sigma\mathbf x^*+\sigma_n^2)$, epistem/ aleator (irr)
$\mathrm{var}[y^*|x^*]=\mathbb E_\theta[\mathrm{var}_{y^*}[y^*|x^*,\theta]]+\mathrm{var}_\theta[\mathbb E_{y^*}[y^*|x^*,\theta]]$
RecUpt: giv. prior $p(\theta)$, obs $y_{1:n}$, $p^{(t)}(\theta)=p(\theta|y_{t+1})$ post. aftr t obs, $p^{(t+1)}(\theta)=p(\theta|y_{1:t+1})=p^{(t)}\cdot p(y_{t+1}|\theta)$, $X_{t+1}^TX_{t+1}=X_t^TX_t+x_{t+1}x_{t+1}^T$
$X_{t+1}^Ty_{t+1}=X^Ty_t+y_{t+1}x_{t+1}$ **Func View:** instead of $w\sim\mathcal N(0,\sigma_p^2I)$, f prior $f|X\sim\mathcal N(\Phi\mathbb E[w],\Phi\,\mathrm{var}[w]\Phi^\top)=\mathcal N(0,\sigma_p^2\Phi\Phi^\top)$,
$k(x,x')=\sigma_p^2\cdot\phi(x)^\top\phi(x')=\mathrm{cov}[f(x),f(x')]$
CV: $\lambda=\hat{\sigma_n^2}/\hat{\sigma_p^2}$, $\hat{\sigma_n^2}$ (MSE), find $\hat{\sigma_p^2}=\hat{\sigma_n^2}/\hat\lambda$
$X_{t+1}\perp X_{1:t-1},Y_{1:t-1}|X_t, Y_t\perp Y_{1:t-1}|X_{t-1}, Y_t\perp X_{1:t-1}|X_t$
State X_t obs Y_t Prior $P(X_1)\sim\mathcal N(\mu,\Sigma)$
Mot: $X_{t+1}=FX_t+\epsilon_t$, $\epsilon_t\sim\mathcal N(0,\Sigma_x)$ **Sens:** $Y_t=HX_t+\eta_t$, $\eta_t\sim\mathcal N(0,\Sigma_y)$ **Kupdt 4 pred:** $\mu_{t+1}=F\mu_t+K_{t+1}(y_{t+1}-HF\mu_t)$ $\Sigma_{t+1}=(I-K_{t+1}H)(F\Sigma_tF^T+\Sigma_x)$ **Kgain:** $K_{t+1}=(F\Sigma_tF^T+\Sigma_x)H^T(H(F\Sigma_tF^T+\Sigma_x)H^T+\Sigma_y)^{-1}$

GP $f\sim GP(\mu(x),K(x)$ (∞-dim Gaussian)
∞ set of RVs X s.t. $\forall A\subseteq X, A=\{x_1,..,x_m\}$ it holds $Y_A=[Y_{x_1},..,Y_{x_m}]\sim\mathcal N(\mu_A,K_{AA})$ w\ $K_{AA}^{(ij)}=k(x_i,x_j)$ and $\mu_A^{(i)}=\mu(x_i)$ (join Gaus)
Kern: *sqrd exp* anal, ∞ dif *exp* cont nodif, *Mat(ν)* ν mal dif, $\nu=1/2$ Lapl, $\nu=\infty$ Gaus
Cov k: symmetric, PSD, *statry:* $k(x,x')=k(x-x')$, *isotropic:* $k(x,x')=k(||x-x'||_2)$.
Pred $p(f)=GP(f;\mu(x),k(x,x'))$, observe $y_i=f(x_i)+\epsilon_i$, $\epsilon_i\sim\mathcal N(0,\sigma^2)$, $A=\{x_1,..,x_m\}$. Often $\mu(x)=0$, $p(f|x_{1:m},y_{1:m})=GP(f;\mu',k')$ with $k_{x,A}=[k(x,x_1)..k(x,x_m)]$, $\mu'(x)=\mu(x)+\mathbf k_{x,A}(\mathbf K_{AA}+\sigma^2\mathbf I)^{-1}(\mathbf y_A-\mu_A)$ $k'(x,x')=k(x,x')-\mathbf k_{x,A}(\mathbf K_{AA}+\sigma^2\mathbf I)^{-1}\mathbf k_{x',A}^T$, *Pred post:* at $x^\star$, $p(y^*|x_{1:m},y_{1:m},x^*)=\mathcal N(\mu_n^*,\sigma_n^{2*})$, $\mu_n^*=\mu'(x^*)$, $\sigma_n^{2*}=\sigma^2+k'(x^*,x^*)$, **Samp:** 1) discrete set to sample $\mathbf f=[f_1...f_n]$, $\mathbf f\sim\mathbf K^{1/2}\epsilon=\mathbf L\epsilon$ 2) $p(f_1..f_n)=\prod_{i=1}^n p(f_i|f_{1:i-1})$, sample $f_n\sim p(f_n|f_{1:n-1})$ **ModelSel:** 1) MLE $\hat\theta=\mathrm{amax}_\theta\, p(y|X,\theta)$, $\int p(y|X,f)p(f|\theta)df$, $\exp\{-\frac12\mathbf y^T\mathbf K_y^{-1}\mathbf y-\frac12\log|\mathbf K_y|\}$ 2) place hyper-prior $p(\theta)$, MAP $\hat\theta=\mathrm{amax}_\theta\, p(y|x,\theta)p(\theta)$, FullB $p(y^\star|x^\star,x_{1:n},y_{1:n})=\int p(y^\star|x^\star,f)p(f|x,y,\theta)p(\theta)d\theta$
Cost: n variables, req. lin system time $\mathcal O(n^3)$, space $\mathcal O(n^2)$, **BLR:** $\mathcal O(nd^2)$. **Fast GP:** 1) *GPU* 2) *Local:* distance decaying kernel (e.g. RBF), only condition on points x' where $|k(x,x')|>\tau$ 3)a) k approx: $k(x,x')\approx\phi(x)^T\phi(x')$, $\phi\in\mathbb R^m$, then BLR $\mathcal O(nm^2+m^3)$ 3)b) stat kern→ *Bochner*, **Random FF** $k_{x,x'}=\int_{\mathbb R^d}p(\omega)e^{j\omega^T(x-x')}d\omega=\mathbb E_{\omega,b}[z_{\omega,b}(x)z_{\omega,b}(x')]\approx\frac1m\sum_i z_{w^{(i)},b^{(i)}}(x)z_{w^{(i)},b^{(i)}}(x')$, $\omega\sim p(\omega)$, $b\sim\mathcal U[0,2\pi]$, $z_{\omega,b}(x)=(2/D)^{1/2}\cos(\omega^Tx+b)$ → draw samples ω_i,b_i, $k(x,x')\approx\phi(x)^T\phi(x')$ with $\phi_i(x)=\frac{1}{\sqrt m}z_{w^{(i)},b^{(i)}}(x)$
Rahimi: $M\subset\mathbb R^d$ compact, for RFFs $z(x)$, $\sigma_p^2=\mathbb E[\omega^T\omega]$, $\mathbb P[\sup_{x,x'\in M}|z(x)^\top z(x')-k(x,x')|\ge\epsilon]\le 2^8(\frac{\sigma_p\,\mathrm{diam}(M)}{\epsilon})^2\exp(-\frac{D\epsilon^2}{4(d+2)})$
4) *Inducing Points methods:* Summarz data via values of f at inducing points $\mathbf u=[u_1,..,u_m]$.
$p(\mathbf f^*,\mathbf f)=\int p(\mathbf f^*,\mathbf f,\mathbf u)d\mathbf u=\int p(\mathbf f^*,\mathbf f|\mathbf u)p(\mathbf u)d\mathbf u$
$p(\mathbf f^*,\mathbf f)\approx q(\mathbf f^*,\mathbf f)=\int q(\mathbf f^*|\mathbf u)q(\mathbf f|\mathbf u)p(\mathbf u)d\mathbf u$ with $u=f(z)$, z-inducing location, train conditional: $p(\boldsymbol f|\mathbf u)=\mathcal N(\boldsymbol K_{\mathbf f,\mathbf u}\boldsymbol K_{\mathbf u,\mathbf u}^{-1}\mathbf u,\boldsymbol K_{\mathbf f,\mathbf f}-\boldsymbol Q_{\mathbf f,\mathbf f})$, w\ $\mathbf Q_{\mathbf a,\mathbf b}=\mathbf K_{\mathbf a,\mathbf u}\mathbf K_{\mathbf u,\mathbf u}^{-1}\mathbf K_{\mathbf u,\mathbf b}$, test cond: $p(\boldsymbol f^*|\mathbf u)=\mathcal N(\boldsymbol K_{\mathbf f^*,\mathbf u}\boldsymbol K_{\mathbf u,\mathbf u}^{-1}\mathbf u,\boldsymbol K_{\mathbf f^*,\mathbf f^*}-\boldsymbol Q_{\mathbf f^*,\mathbf f^*})$
4)a) *Subset of Regressors:* assume $\mathbf K_{\mathbf f,\mathbf f}-\mathbf Q_{\mathbf f,\mathbf f}=0$, replace $p(\mathbf f|\mathbf u)$ by $q_{SoR}(\mathbf f|\mathbf u)=\mathcal N(\mathbf K_{\mathbf f,\mathbf u}\mathbf K_{\mathbf u,\mathbf u}^{-1}\mathbf u,0)$ resulting model is degenerate GP with covariance function $k_{SoR}(\mathbf x,\mathbf x')=k(x,\mathbf u)\mathbf K_{\mathbf u,\mathbf u}^{-1}k(\mathbf u,\mathbf x')$ *FITC:* Assume $\mathbf f_i\perp\mathbf f_j|\mathbf u,\forall i\ne j$ $q_{FITC}(\mathbf f|\mathbf u)=\mathcal N(\mathbf K_{\mathbf f,\mathbf u}\mathbf K_{\mathbf u,\mathbf u}^{-1}\mathbf u,\mathrm{diag}(\mathbf K_{\mathbf f,\mathbf f}-\mathbf Q_{\mathbf f,\mathbf f}))$, $q_{FITC}(\boldsymbol f^*|\mathbf u)=\mathrm p(\boldsymbol f^*|\mathbf u)$, cost cubic in # inducing pts, dominated inv $K_{\mathbf u,\mathbf u}$. *Pick inducing pts* by chose randly/greed criterion (var)/det grid, $\mathbf u$ as hyperpar, max marg like (ensure $\mathbf u$ repr data).

Variatnl Inf (rev greed sel mode, fwd sel var)
$\mathrm{amin}_\lambda KL(p\|q_\lambda)=\mathrm{amax}_\lambda\lim_{n\to\infty}\sum_{i=1}^n\log q(x^{(i)}|\lambda)$
$q=\arg\min_{q\in\mathcal Q}KL(p||q)$ match 1, 2 mom of p
Laplace: $p(\theta|(x,y)_{1:n})\approx q_\lambda(\theta)=\mathcal N(\hat\theta,\Lambda^{-1})$
$\hat\theta=\arg\max_\theta p(\theta|y)$, $\Lambda=-\nabla^2\log p(\hat\theta|y),0=\nabla\log p(\hat\theta|y)$, *Pred* BL, greed fit mode (overcnf), match curv, pres MAP est, can diff from post
GVI: $0=\mathbb E_{q(\boldsymbol\theta)}[\nabla_{\boldsymbol\theta}\log p(\mathcal D,\boldsymbol\theta)]$, $\boldsymbol\Sigma^{-1}=-\mathbb E_{q(\boldsymbol\theta)}[\nabla^2_{\boldsymbol\theta}\log p(\mathcal D,\boldsymbol\theta)]$ **PVI:** $KL(\mathcal N\|Po)\propto -d/2\ln(2\pi e)$-$d\ln\sigma$-$1/m\sum_{j,i=1}^{m,n}[y_i\boldsymbol w^\top\boldsymbol x_i$-$e^{\boldsymbol w^\top\boldsymbol x_i}$-$\|\boldsymbol w\|_2^2/2\sigma_p^2]_{\boldsymbol w=\boldsymbol\mu+\sigma\epsilon^{(j)}}$ **BLR:** $P(y|x,w)=\mathrm{Ber}(y;\sigma(w^Tx))$, $\hat w=\mathrm{amin}_w\sum_{i=1}^n\log(1+\exp(-y_iw^Tx_i))+\lambda\|w\|_2^2$, $\mathcal N(\hat w,\Lambda^{-1})$, $\Lambda=\sum_{i=1}^n x_ix_i^T\sigma(\hat w^Tx_i)(1-\sigma(\hat w^Tx_i))$
Variational Inf: $p(\theta|y)=\frac1Z p(\theta,y)\approx q_\lambda(\theta)$
$q^*\in\arg\min_{q\in\mathcal Q}KL(q||p)$: $q\approx p$ where q large
$\mathrm{amin}_q KL(q||p)=\mathrm{amax}_q\mathbb E_{\theta\sim q}[\log p(\theta,y)]+H(q(\theta))$ $=\mathrm{amax}_q\mathbb E_{\theta\sim q_\lambda(\theta)}[\log p(y|\theta)]-KL(q(\theta)||p(\theta))$
$\max_\epsilon\mathbb E_\epsilon[\ln p(\mathcal D,\mu+\Sigma^{1/2}\epsilon)]+\frac12\ln|\Sigma|+\frac D2\log 2\pi e$
ELBO, $L(\lambda)\le\log p(y)$. $\nabla_\lambda L(\lambda)$ hard, *score* ∇: $\nabla_\lambda L=\mathbb E_{\theta\sim q_\lambda}[\nabla_\lambda\log q(\theta|\lambda)(\log p(y,\theta)$-$\log q(\theta|\lambda))]$ $\theta\sim q_\lambda(\cdot)$ dep on var param. **Reparam.:** let $\epsilon\sim\phi$, $\theta=g(\epsilon,\lambda)$, $q(\theta|\lambda)=\phi(\epsilon)|\nabla_\epsilon g(\epsilon;\lambda)|^{-1}$ and $\mathbb E_{\theta\sim q_\lambda}[f(\theta)]=\mathbb E_{\epsilon\sim\phi}[f(g(\epsilon;\lambda))]$, which yield $\nabla_\lambda\mathbb E_{\theta\sim q_\lambda}[f(\theta)]=\mathbb E_{\epsilon\sim\phi}[\nabla_\lambda f(g(\epsilon;\lambda))]$ *Blackbox VI:* max ELBO using stoch opt., for diagonal q, 2x expensive as MAP, only need to diff joint prob p and q, also use Natural Grad, Variance reduct tech. **GP Class:** $P(f)=GP(\mu,k)$, $P(y|f,\mathbf x)=\sigma(y\cdot f(\mathbf x))$ max (*) w\ $q(f_i):=\int p(f_i|\mathbf u)q(\mathbf u)d\mathbf u$, $\mathbf u$ *pseudo in.* $\sum_{i=1}^n\mathbb E_{q(f_i)}[\log p(y_i|f_i)]-KL(q(\mathbf u)\|p(\mathbf u))$

MCMC (MC: seq of RV $X_{1:n}$ with (*))
TV Dist: $\|\mu-v\|_{\mathrm{TV}}=2\sup_{A\subseteq\mathcal A}|\mu(A)-v(A)|$
Mix.Time: $\tau_{\mathrm{TV}}(\epsilon)=\min\{t|\forall q_0:\|q_t-\pi\|_{\mathrm{TV}}\le\epsilon\}$
Rapidly mix: $\tau_{\mathrm{TV}}(\epsilon)\in\mathcal O(\mathrm{poly}(n,\log(1/\epsilon)))$
(*) Stat MC w\ prior $P(X_1)$, trans $P(X_{t+1}|X_t)$ indep of t. *ergodic* $\exists\, t<\infty$ s.t. all states reachable from every state in *exactly* t steps. **Mark Ass:** $X_{t+1}\perp\!\!\!\perp X_{1:t-1}|X_t\forall t$ **Stat Distrib:** $\pi(P-I)=0$. Stat Ergodic MC is unique stat distr $\pi(X)>0$ s.t. $\forall x$: $\lim_{N\to\infty}P(X_N=x)=\pi(x)$, $\pi(X)$ indep of prior $P(X_1)$. **Sim MC** fwd sampl $x_N\sim P(X_N|X_{N-1}=x_{N-1})$ **MCMC:** Approx pred. distr. $p(y^*|x^*,x_{1:n},y_{1:n})=\int p(y^*|x^*,\theta)p(\theta|(x,y)_{1:n})d\theta=\mathbb E_{\theta\sim p(\cdot|(x,y)_{1:n})}[p(y^*|x^*,\theta)]\approx\frac1m\sum_{i=1}^m p(y^*|x^*,\theta^{(i)})$, sample $\theta^{(i)}\sim p(\theta|(x,y)_{1:n})$ from MC with stat distr (*).
Hoeffding: Let f be bounded in $[0,C]$:
$\mathbb P(|\mathbb E[f(X)]$- $\frac1N\sum_1^N f(x_i)|>\epsilon)\le 2\exp(\text{-}2N\epsilon^2/C^2)$
Given unnormalized distr. $Q(x)>0$, design MC s.t. $\pi(x)=\frac1ZQ(x)$. **DB (reversible):** $Q(x)P(x'|x)=Q(x')P(x|x')\to\pi(x)=\frac1ZQ(x)$.

Gibbs Sampling: Asympt. correct but slow
1. Init $\mathbf x^{(0)}$, fix observed RVs X_B to $\mathbf x_B$
2. Repeat: set $\mathbf x^{(t)}=\mathbf x^{(t-1)}$; sel $j\in[1:m]\setminus B$ $x_j^{(t)}\sim P(X_j|\mathbf x^{(t)}_{[1:m]\setminus\{j\}})$ Rand: ful. DB, find cor distr. Determin: not ful. DB, corr distr.

$P(X_i=x_i|\boldsymbol x_{-i})=\frac1ZQ(X_i=x_i,\boldsymbol x_{-i})=\frac1ZQ(\boldsymbol x_{1:n})$ re-sampl X_i only requires eval unnorm joint distrib and renorm. **Expectations via MCMC**: Joint sample at t dep only on t-1 → LLN, HB not apply. *Thm:* $X_{1:n}$ EMC on finite D, $f\in D$, $\lim_{N\to\infty}\frac1N\sum_{i=1}^N f(x_i)=\sum_{x\in D}\pi(x)f(x)$
Use MCMC to get samples $\mathbf X^{(1:T)}$. After burn-in time t_0: $\mathbb E[f(\mathbf X)]\approx\frac{1}{T-t_0}\sum_{\tau=t_0+1}^T f(\mathbf X^{(\tau)})$
MetroH: Gen MC → DB 1) Proposal $R(X'|X)$, given $X_t=x$, sample $x'\sim R(X'|X=x)$; 2) For $X_t=x$, w.p. $\alpha=\min\{1,\frac{Q(x')R(x|x')}{Q(x)R(x'|x)}\}$: $X_{t+1}=x'$; else $X_{t+1}=x$ **Cont RVs:** log-concave $p(x)=\frac1Z\exp(-f(x))$, f convex. M/H: $\alpha=\min\{1,\frac{R(x|x')}{R(x'|x)}\exp(f(x)-f(x'))\}$, Simple uninf dir. **Improved prop:** *MALA/LMC:* $R(x'|x)=\mathcal N(x';x-\eta_t\nabla f(x);2\tau I)$ MALA converg to stat distr for log concave (locall also non convex) func (for general distrib convergence slow), mixing time $\mathcal O(d)$. *Improve efficiency:* both proposal step and accept step requires full acces to energy function f → SGD, decaying step size, skip accept/reject → *SGLD*
$\theta\sim\frac1Z\exp(\log p(\theta)+\sum_{i=1}^n\log p(y_i|x_i,\theta))$ 1) θ_0 2) For t do $i_1..i_m\sim U(1,n)$, $\epsilon_t\sim\mathcal N(0,2\eta_tI)$, $\theta_{t+1}=\theta_t$-$\eta_t(\nabla\ln p_{\theta_t}+\frac nm\sum_1^m\nabla\ln p(y_{i_j}|\theta_t,x_{i_j}))+\epsilon_t$ SGLD = SGD + Gauss noise, convergence if $\eta_t\in\mathcal O(t^{-1/3})$, const step boost mixing, improve perf. via Adagrad, *HMC* (mom)

BDL (Prior: $p(\theta)=\mathcal N(\theta;0,\sigma_p^2I)$, Gauss = *)
* weig decay, *Like:* $p(y|x,\theta)=\mathcal N(y;f(x,\theta),\sigma^2)$ MAP: $\hat\theta=\mathrm{amin}_\theta-\log p(\theta)-\sum_i\log p(y_i|x_i,\theta)$ hetero ϵ well, fails pred epistemic, use *VI (BbB)*. **Etero:** $y|\boldsymbol x,\boldsymbol\theta\sim\mathcal N(\mu(\boldsymbol x;\boldsymbol\theta),\sigma^2(\boldsymbol x;\boldsymbol\theta))$, $f_1(\boldsymbol x;\boldsymbol\theta)$, $\exp(f_2(\boldsymbol x;\boldsymbol\theta))$ **VI:** SGD-opt ELBO via $\nabla_\lambda L(\lambda)$. Find VI approx q_λ. Draw m weights $\theta^{(j)}\sim q_\lambda(\cdot)$. *Pred* $p(y^*|\mathbf x^*,\mathbf x_{1:n},y_{1:n})\approx\mathbb E_{\theta\sim q_\lambda}[p(y^*|\mathbf x^*,\theta)]\approx\frac1m\sum_j p(y^*|\mathbf x^*,\theta^{(j)})$, $\mathrm{Var}[y^*|\cdot]\approx\frac1m\sum_1^m\sigma^2(x^*,\theta^{(j)})+\frac1m\sum_1^m(\mu(x^*,\theta^{(j)})-\overline\mu(x^*))^2$
MC wghts $\theta^{(1:T)}$ SGLD, LD, SGHMC; pred by avg $\theta^{(1:T)}$. *Summ:* subsmpl/GVI $q(\theta|\mu_{1:d},\sigma^2_{1:d})$ $\mu_i=\frac1T\sum_{j=1}^T\theta_i^{(j)}$; $\sigma_i^2=\frac1T\sum_{j=1}^T(\theta_i^{(j)}-\mu_i)^2$
DropVI: $q_j(\theta_j|\lambda_j)=p\delta_0(\theta_j)+(1\text{-}p)\delta_{\lambda_j}(\theta_j)$, $\theta^{(j)}$ NN w\ wght giv by λ, $\theta^{(j)}=0$ wp p.
Prob. Ensbles: train multiple models bootstrap, **Cal:** confidence=accuracy on heldout
Reliability Diag: plot expected sample acc

as func of confidence 1) group pred in M bin, 2) freq$(B_m) = 1/|B_m| \sum_{i\in B_m} [\hat y_i = 1]$ 3) conf$(B_m) = 1/|B_m|\sum_{i\in B_m}\hat p_i$ 4) $MECE = \max \sum_{m=1}^M \frac{|B_m|}{n}|\text{freq}(B_m) - \text{conf}(B_m)|$ *Improve* via histo bin, isotonic regr., platt (temp) scaling

Active Learn (min $\#x$ reducing uncertainty)

MutInf: $I(X;Y) = I(Y;X) = H(X)$-$H(X|Y)$ $X\sim N(\mu,\Sigma), Y = X + N(0,\sigma^2)$, $I = \frac12 \ln|I + \frac{1}{\sigma^2}\Sigma|$

InfGain: $f(S)$, $S\subseteq D$, $F(S) = H(f) - H(f|y_S) = I(f;y_S) = \frac12\log|I+\sigma^{-2}K_S|$, obs at S
Thm: $F(S_T) \ge (1-1/e)\max_{S\subseteq D,|S|\le T} F(S)$

Greed MIopt: $F(S)$ NP hard, $S_t = \{x_1,..,x_t\}$ $x_{t+1} = \arg\max_{x\in D} F(S_t\cup\{x\})$ =arg max$_{x\in D}\sigma^2_{x|S_t}$

UncSmpl: $x_t = \arg\max_{x\in D}\sigma^2_{t-1}(x)$, (homo, $\mathcal N$) Fail distepist/aleat, *Etero:* amax$_x$ $\sigma_f^2(x)/\sigma_n^2(x)$
Or: trAce/Eigval/log-Determinant designs

BALD: (class) $x_{t+1} = \text{amax}_x I(\theta; y_x|x_{1:t}, y_{1:t}) = \text{amax}_x H(y|x,(x,y)_{1:t}) - \mathbb E_{\theta\sim p(\cdot|(x,y)_{1:t})}[H(y|x,\theta)]$

Bayesian Opt (seq. pick $x_1,..,x_T\in D$ (*))

(*), $y_t = f(x_t)+\epsilon_t$, find $\max_x f(x)$ st T smal

CumReg: $R_T = \sum_{t=1}^T \max_{x\in D} f(x) - f(x_t)$ sublin if $R_T/T\to 0$ (obj) $\Leftrightarrow f(x_t)\to\max f(x)$ (UCB $\ge$ best lower bound if well cal) **GP-UCB:** $x_t = \arg\max_{x\in D}\mu_{t-1}(x) + \beta_t\sigma_{t-1}(x)$ $\mu(x),\sigma(x)$ from GP marginal, β_t EE-tradeoff. non-convex usually, low D use Lipschitz, high D. grad ascent (w\ rand init) *Thm:* $f\sim GP$,if correct β_t: $\frac1T R_T = \mathcal O(\sqrt{\gamma_T/T})$, with $\gamma_T = \max_{|S|\le T} I(f;y_S)$ (max. info gain)

Thomp sampl: at t, draw from GP post. $\tilde f\sim P(f|x_{1:t},y_{1:t})$, select $x_{t+1}\in\arg\max_{x\in D}\tilde f(x)$ datasets in BO/AL small, data pts selected dependend on prior obs, *sol:* hyperprior on hyperparam, select pts at random occasionally.

Markov Decision Processes (r, P known)

MDP: Finite MDP (control MC), state $X = \{1,..,n\}$, action $A = \{1,..,m\}$, trans $P(x'|x,a)$, init $P(x_0)$, reward $r(x,a,x')$, discount $\gamma\in[0,1]$

Planning in MDPs: Policy $\pi: X\to AP(A)$ (det./rand), induce MC w\ trans $P(X_{t+1} = x'|X_t = x) = P(x'|x,\pi(x)) \sum_a \pi(a|x)P(x'|x,a)$

Value fun: fixed π $V^\pi(x) = J(\pi|X_0=x) = \mathbb E[\sum_{t=0}^\infty \gamma^t r(X_t,\pi(X_t))|X_0=x] = r(x,\pi(x)) + \gamma\sum_{x'}P(x'|x,\pi(x))V^\pi(x')$, $V^\pi = (I-\gamma T^\pi)^{-1} r^\pi$ $V_i^\pi = V^\pi(i)$, $r_i^\pi = r^\pi(i,\pi(i))$, $T^\pi_{i,j} = P(j|i,\pi(i))$
$V^\pi(x) = \sum_{x'}P(x'|x,\pi(x))[r(x,\pi(x),x') + \gamma V^\pi(x')]$ $= Q^\pi(x,\pi(x)) = \mathbb E_{a'\sim\pi(x)}Q^\pi(x,a')$ det/rand

Fixed Point Iteration: 1) initialize V_0^π 2) for $t = 1:T$ do: $V_t^\pi = r^\pi + \gamma T^\pi V^\pi_{t-1}$ (converges)

Greedy policy w.r.t. V: V induces policy $\pi_V(x) = \text{amax}_a r(x,a) + \gamma\sum_{x'}P(x'|x,a)V(x')$

Greedy π w.r.t. Q: $\pi_Q(x) = \arg\max_a Q(x,a)$

BellThm: Opt $\pi^*\leftrightarrow$greedy wrt induced V^* $V^*(x) = \max_{a\in\mathcal A}[r(x,a) + \gamma\sum_{x'\in X}P(x'|x,a)V^*(x')]$ $= \text{mx}_{a\in\mathcal A}\mathbb E_{x'}[r(x,a)+\gamma V^*(x')] = \text{mx}_{a\in\mathcal A}Q^*(x,a)$

Policy Iter: 1) Init arbitry π 2) Until conv: calc $V^{\pi_t}(x)$, calc greedy pol π_t^G w.r.t. V^{π_t}, set $\pi_{t+1}\leftarrow\pi_t^G$ Stop if $V^{\pi_t}(x) = V^{\pi_{t+1}}(x)$. Monoton improves all val $V^{\pi_{t+1}}(x)\ge V^{\pi_t}(x)\forall x$. Converge exact optimal in $\mathcal O(n^2 m/(1-\gamma))$. **SAF:** $Q_t(x,a) = r(x,a) + \gamma\sum_{x'}P(x'|x,a)V_{t-1}(x')$

Value Iteration: 1) Init $V_0(x) = \max_a r(x,a)$ 2) for $t = 1:\infty$: $V_t(x) = \max_a Q_t(x,a)$. Stop if $||V_t - V_{t-1}||_\infty\le\epsilon$, pick grdy π_G w.r.t. V_t. Conv ϵ-opt sol in poly time. *Tradeoffs:* Pol/val $\mathcal O(n^3/nmk)$ per iter, k sprs **POMDP:**(X,A,P,R,γ,Y,Q), $O_{x,y,a} = P[Y_{t+1}=y|X_{t+1}=x, A_t = a]$

Belief state: $b_t(x) = P[X_t = x|y_{1:t}, A_{t-1} = a]$ $b_t(x)\in\Delta^{|X|} = \{b\ge0\in\mathbb R^{|X|}, \sum_{i=1}^{|X|}b_i = 1\}$ $b_{t+1}(x) = \frac1Z O(y_{t+1}, x)\Sigma_{x'}b_t(x')P(x|x',a), Z = \sum_x$

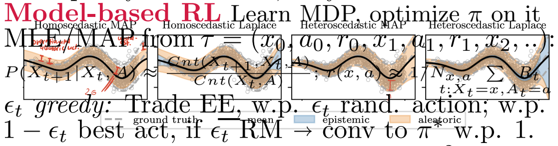
BMDP: $(\Delta^{|X|}, A, \tau(b'|b,a), \rho = \sum_x b(x)r(x,a))$

Solve POMDPs finite T: exp. #belif states. BUT: most belif states never reach → discretize space by sampling, PBVI, PBPI, dim reduction

RL (agent act. chang state, unknown MDP)

- **On**-policy: full control on actions/EE-trade
- **Off**-policy: no control, only observation data

Model-based RL Learn MDP, optimize π on it MLE/MAP from $\tau = (x_0, a_0, r_0, x_1, a_1, r_1, x_2, ..)$: $P(X_{t+1}|X_t,A) \approx \frac{Cnt(X_{t+1},X_t,A)}{Cnt(X_t,A)}$; $r(x,a) \approx 1/N_{x,a}\sum_{t:X_t=x,A_t=a} R_t$



ϵ_t *greedy:* Trade EE, w.p. ϵ_t rand. action; w.p. $1-\epsilon_t$ best act, if ϵ_t RM → conv to π^* w.p. 1.

RobMonro (RM): $\sum_t\epsilon_t = \infty$, $\sum_t\epsilon_t^2<\infty$

$R_{\max}$ Alg: Set unknown $r(x,a)$ to R_{max}, $r(x,a)\le R_{max},\forall x,a$, add fairy tale state x^*, set $P(x^*|x,a) = 1$, compute π. Repeat: run π while updtng $r(x,a)$, $P(x'|x,a)$, recompute π. $P(x^*|x^*,a) = 1, r(x^*,a) = R_{\max}$ *Lem:* evry T stps, whp $R_{\max}$ eithr obt near opt reward, or visit at least 1 unknwn state-action pair. *Thm:* wp 1-δ, $R_{\max}$ reach ϵ-opt policy in #steps poly in $|X|,|A|,T,1/\epsilon,\log(1/\delta),R_{\max}$. **Problms of MBRL:** *Mem require:* $P(x'|x,a)\approx\mathcal O(|X|^2|A|)$, $r(x,a)\approx\mathcal O(|X||A|)$ *Computate:* reptdly solve MDP (Val/Pol Iter)

Model-free RL Directly estim value function

TD-Learning: Follow policy π, get (x,a,r,x') Update: $\hat V^\pi(x)\leftarrow(1-\alpha_t)\hat V^\pi(x) + \alpha_t(r+\gamma\hat V^\pi(x'))$ *Thm:* If α_t is RM and all (x,a) pairs chosen ∞ often, then $\hat V^\pi$ converges to V^π w.p. 1.

Optimistic Q-L Estimate $Q^*(x,a)$ 1) Init rand /zero $\hat Q^*(x,a) = \frac{R_{max}}{1-\gamma}\prod_{t=1}^{T_{init}}(1-\alpha_t)^{-1}$ 2) at t $a_t\in\text{amax}_a\hat Q^*(x_t,a)$, get (x_t,a_t,r,x'), $\hat Q^*(x_t,a_t)$ ← (1-α_t)$\hat Q^*(x_t,a_t) + \alpha_t(r+\gamma\max_{a'}\hat Q^*(x',a'))$
Thms for $R_{\max}$, TD-L → Time: $\mathcal O(|A|)$, Mem: $\mathcal O(|X||A|)$, $|X|,|A|$, exp #agents, state vars

RL via FuncAprox (parametric approx. of (*))

(*) val funct $V(x;\theta)$ or act val funct $Q(x,a;\theta)$

TD-Learning as SGD: Tabular TD update rule can be viewed as SGD on squared loss $l_2(\theta;x,x',r) = \frac12(V(x;\theta) - r - \gamma V(x';\theta_{old}))^2$, then $V\leftarrow V - \alpha_t\nabla_{V(x;\theta)}l_2$ is equiv to TD update.

Function Approx Q-learning: very slow Loss $l_2(\theta;x,a,r,x') = \frac12\delta^2$ with $\delta = Q(x,a;\theta) - r - \gamma\max_{a'}Q(x',a';\theta_{old})$. *Alg:* Until conv: In state x, pick action a, obsrv r,x'. Update: $\theta\leftarrow\theta - \alpha_t\nabla_\theta l_2 = \theta - \alpha_t\delta\nabla_\theta Q(x,a;\theta)$ **DQN:** Q-learn w\ NN as func aprox, *experience replay*, maintn data D, clone NN to stablze target opt $L(\theta) = \sum_{(x,a,r,x')\in D}(r+\gamma\max_{a'}Q(x',a';\theta^{old}) - Q(x,a;\theta))^2$

Double DQN: Use current NN to eval action arg max to prevents maximiz. bias of DQN. $L^{DDQN}(\theta) = \sum_{(x,a,r,x')\in D}[r+\gamma Q(x',a^*(\theta);\theta^{old}) - Q(x,a;\theta)]^2$ $a^*(\theta) = \arg\max_{a'}Q(x',a';\theta)$

Q-learn π $a_t = \text{amax}_a Q(x,a;\theta)$ bad if $|A|\uparrow$

Policy Search Methods: Param. policy π_θ Max $J(\theta) = \mathbb E_{\tau\sim\pi_\theta}[r(\tau)]$ $(\tau = x_{0:T}, a_{0:T}\sim\pi_\theta)$, $r(\tau) = \sum_{t=0}^T\gamma^t r(x_t,a_t)$); via ∇_θ. *Score Grad:* $\nabla_\theta J_\theta = \nabla_\theta\mathbb E_{\tau\sim\pi_\theta}r(\tau) = \mathbb E_{\tau\sim\pi_\theta}[r(\tau)\nabla_\theta\ln\pi_\theta(\tau)]$ *MDP:* $\pi_\theta(\tau) = p(x_0)\prod_0^T\pi(a_t|x_t;\theta)p(x_{t+1}|x_t,a_t)$ then: $\nabla_\theta J(\theta) = \mathbb E_{\tau\sim\pi_\theta}[r(\tau)\sum_{t=0}^T\nabla_\theta\log\pi(a_t|x_t;\theta)]$ *Reduce variance:* baselines (but ∇ unbiased) $\mathbb E_{\tau\sim\pi_\theta}[r(\tau)\nabla\log\pi_\theta(\tau)] = \mathbb E...[(r(\tau)-b)\nabla\log\pi_\theta(\tau)]$

R2Go: $G_t = \sum_{t'=t}^T\gamma^{t'-t}r_{t'}$; $b_t(x_t) = 1/T\sum_{t=0}^T G_t$ $\nabla J_T(\theta) = \mathbb E_{\tau\sim\pi_\theta}[\sum_{t=0}^T\gamma^t G_t\nabla_\theta\log\pi(a_t|x_t;\theta)]$ *Mean over returns:* replace G_t with $(G_t - b_t(x_t))$

REINFORCE: Input $\pi(a|x;\theta)$, init θ, repeat: gener episode $(x_i,a_i,r_i), i = 0:T$; for $t = 0:T$ set G_t to retrn from step t, update θ: $\theta = \theta + \eta\gamma^t G_t\nabla_\theta\log\pi(A_t|X_t;\theta)$ optimizes score-∇ using MC returns; high variance

Deep RL with policy grad and actor-critic

Advantge Func: $A^\pi(x,a) = Q^\pi(x,a) - V^\pi(x)$ π^*iff $\forall x,a: A^{\pi^*}(x,a)\le0; \forall\pi,x:\max_a A^\pi(x,a)\ge0$

Actor-Critic: Approx π_θ and V^π, e.g. 2 NNs Reinterpreting score-∇ = *Policy grad Thm* $\nabla J(\theta_\pi) = \mathbb E_{\tau\sim\pi_\theta}[\sum_{t=0}^\infty\gamma^t Q(x_t,a_t;\theta_Q)\nabla\log\pi(a_t|x_t;\theta_\pi)]$ $= \mathbb E_{(x,a)\sim\pi_\theta}[Q(x,a;\theta_Q)\nabla_{\theta_\pi}\log\pi(a|x;\theta_\pi)]$;

Online A.C.: val fun apprx + pol grad thm $\delta = Q(x,a;\theta_Q) - r - \gamma Q(x',\pi(x',\theta_\pi);\theta_Q)$ $\theta_\pi\leftarrow\theta_\pi + \eta_t Q(x,a;\theta_Q)\nabla\log\pi(a|x;\theta_\pi)$ $\theta_Q\leftarrow\theta_Q - \eta_t\delta\nabla Q(x,a;\theta_Q)$ (**FA Q-learning**) *Var red:* baselines $Q(x,a;\theta_Q) - V(x;\theta_V)$: adv func estim → **A2/3C, GAE/GAAC**; **TRPO** opt surrogate obj in trust region, guarantees monoton improve $J(\theta)$, **PPO** heuristic variant

Off-policy AC: allow reuse of past data Replace in L_{DQN} a' by $\pi(x';\theta_\pi)$, π follows greedy pol., to model $\max_{a'}Q$. Equivalent to: $\theta_\pi^*\in\arg\max_\theta\mathbb E_{x\sim\mu}[Q(x,\pi(x;\theta);\theta_Q)]$, with $\mu(x)>0$ *explores all states.* If $Q(\cdot;\theta_Q),\pi(\cdot;\theta_\pi)$ diff, use backprop to get stoch-∇ (unbiased) $\nabla_\theta J(\theta) = \mathbb E_{x\sim\mu}[\nabla_\theta Q(x,\pi(x;\theta);\theta_Q)]$ $\nabla_\theta Q(x,\pi(x;\theta)) = \nabla_a Q(x,a)|_{a=\pi(x;\theta)}\nabla_\theta\pi(x;\theta)$ Needs *deterministic* π → inject additional action noise (ϵ_t greedy) to ensure explore →

DDPG: 1) init θ_Q,θ_π, set $\theta_Q^{old} = \theta_Q, \theta_\pi^{old} = \theta_\pi$ 2) repeat: observe x, execute $a = \pi(x;\theta_\pi)+\epsilon$ to observe r,x', store in D. If time to update: for iter: sample batch B from D, cmpt target $y = r+\gamma Q(x',\pi(x',\theta_\pi^{old}),\theta_Q^{old})$, updates: *Critic:* $\theta_Q\leftarrow\theta_Q - \eta\nabla1/|B|\sum_B(Q(x,a;\theta_Q)-y)^2$, *Actor:* $\theta_\pi\leftarrow\theta_\pi + \eta\nabla1/|B|\sum_B Q(x,\pi(x;\theta_\pi);\theta_Q)$, *Params:* $\theta_j^{old}\leftarrow(1-\rho)\theta_j^{old} + \rho\theta_j$ for $j\in\{\pi,Q\}$

TD3: DDPG with 2 Critics avoid maxim bias

Rand π DDPG: (DQN with reparam π grad) For *critic* $a'\sim\pi(x';\theta_\pi^{old})$, get unbias ∇, *actor* $\nabla_{\theta_\pi}\mathbb E_{a\sim\pi(x;\theta_\pi)}Q(x,a;\theta_Q)$reparam $a = \psi(x;\theta_\pi,\epsilon)$ $\nabla_{\theta_\pi}\mathbb E_{a\sim\pi_{\theta_\pi}}Q(x,a;\theta_Q) = \mathbb E_\epsilon\nabla_* Q(x,\psi(x;\theta_\pi,\epsilon);\theta_Q)$

EntrpyReg: $J_\lambda(\theta) = J(\theta) + \lambda H(\pi_\theta)\uparrow$ explor

SAC: variant of DDPG/TD3 for H reg MDPs

MB DipRL (approx dynamcs model $f\approx p$, r)

Init π, data (or {}), for epis: 1) use π, get data 2) lern f, r from data, 3) plan new π on estim

Plang: cont full obs X, nonlin trans, constr

DetDyn: $x_{t+1} = f(x_t,a_t)$, finit H, at t mxze: $J_H(a_{t:t+H-1}) = \sum_{\tau=t:t+H-1}\gamma^{\tau-t}r_\tau(x_\tau(a_{t:\tau-1}),a_\tau)$ $x_\tau(a_{t:\tau-1}) = f(f(...(f(x_t,a_t),a_{t+1})..))$ do a_t, replan. Opt via ∇ meth for diff r,f, cont A (local min, vanish/explod ∇) → use rand shoot. *Rand shot:* Gen rand $a_{(i),t:t+H-1}$, pick $i^* = \text{amax}_i J_H(a_{(i),t:t+H-1})$. If fin H, sparse r: *MPC ValEstim:* $J_H(a_{t:t+H-1}) = \sum_{\tau=t:t+H-1}\gamma^{\tau-t}r_\tau(x_\tau(a_{t:\tau-1}),a_\tau) + \gamma^H V(x_{t+H})$, **StocDyn:** $\max_{a_{t:t+H-1}}\mathbb E_{x_{t+1:t+H}}[\sum_{\tau=t:t+H-1}\gamma^{\tau-t}r_\tau + \gamma^H V(x_{t+H})|*]$

$\mathbb E$ via *MC traj smplng*, unbias estm of J_H, aprox via smpl avg. **Param π:** ($H=0\to$ DDPG) $J_H(\theta) = \mathbb E_{x_0\sim\mu}[\sum_{\tau=0:H-1}\gamma^\tau r_\tau + \gamma^H Q(x_H,\pi(x_H,\theta))|\theta]$

UnknownDyn: follow π, learn f,r,Q off-π from replay buf, replan π based on f,r,Q. Point est poor perf, err compound → use *BayL:* Model distrib over f (BNN, GP), use aprox inference (exact, VI, MC..) **Greedy exploi:** 1) $D = \{\}$, prior $P(f|\{\})$ 2) repeat: plan new π to $\max_\pi\mathbb E_{f\sim P(\cdot|D)}J(\pi,f)$, use π, add new data to D, update post $P(f|D)$ **PETS algo:** ensmbl of NNs to pred cond Gaus trans distr, MPC for plang. **Explor:** add noise/*/* **Thom Smplng:** Like Greedy but in 2) smpl model $f\sim P(\cdot|D)$ then $max_\pi J(\pi,f)$ Use epist noise, ↑ explor. **Opt explor:** Like Greedy but in 2) $\max_\pi\max_{f\in M(D)}J(\pi,f)$; w\ $M(D)$ set of plausible models given D.