Lecture 17 - MaxEnt duality; exp. family

Monday, November 11, 2024 12:33 PM

today: · max Ent & duality
· exponential family

MLE & KL minimization

$\{p_\theta\}_{\theta\in\Theta}$ parametric family for a discrete observation space

$$\boxed{\text{then ML for } \theta \text{ for iid data} \iff \min_{\theta\in\Theta} KL(\hat{p}_n \,\|\, p_\theta)}$$

$\hat{p}_n(x) \triangleq \frac{1}{n}\sum_{i=1}^{n} \delta(x, x^{(i)})$

"empirical distribution" ; δ: kronecker-delta

proof:

$$KL(\hat{p}_n \,\|\, p_\theta) = \sum_x \hat{p}_n(x)\log\frac{\hat{p}_n(x)}{p_\theta(x)}$$

$$= -H(\hat{p}_n) - \sum_x \underbrace{\hat{p}_n(x)}_{\frac{1}{n}\sum_i \delta(x,x^{(i)})}\log p_\theta(x)$$

$$\frac{1}{n}\sum_{i=1}^{n}\underbrace{\sum_x \delta(x,x^{(i)})\log p_\theta(x)}_{\log p_\theta(x^{(i)})}$$

$$= \underbrace{-H(\hat{p}_n)}_{\text{constant w.r.to } \theta} - \frac{1}{n}\underbrace{\sum_{i=1}^{n}\log p_\theta(x^{(i)})}_{\log\left(\prod_{i=1}^{n} p_\theta(x^{(i)})\right)} \quad //$$

Maximum entropy principle

idea: consider some subset of dist over $\mathcal{X}$ according to some data-driven constraints

get a subset $\mathcal{M} \subseteq \Delta_{|\mathcal{X}|}$ ← prob. simplex over $|\mathcal{X}| = K$ elements

MAXENT principle: pick $\hat{p}\in\mathcal{M}$ which maximizes the entropy

i.e. $$\boxed{\hat{p} = \arg\max_{q\in\mathcal{M}} H(q)}$$

$$= \arg\min_{q\in\mathcal{M}} KL(q \,\|\, \underset{u}{\text{uniform}})$$

constant

$$KL(q\|u) = \sum_x q(x) \log \frac{q(x)}{u(x)} = -H(q) + \log(K)$$ with $u(x) = 1/K$

"generalized max. entropy" $KL(q\|h_0)$ ← preferred dist. to bias towards

* example from Wainwright

$\hat{p}_L = \frac{3}{4}$ kangaroos are left-handed

$\hat{p}_B = \frac{2}{3}$ " drink Sobatt beer

question: how many kangs are both L.H. & drink Sab. beer

[here: max. entropy solution is that $P(B=1, LH=1) = \hat{p}_B \cdot \hat{p}_L$ (indep)]

* how do we get set M?

typically: through empirical "moments"

kangaroo: $T_1(x) = \mathbb{1}\{X \text{ drinks beer}\}$
$T_2(x) = \mathbb{1}\{X \text{ is L.H.}\}$

feature functions: $T_1(x), T_2(x), \dots, T_d(x)$ d features

define $M = \{ q : \underbrace{\mathbb{E}_q[T_j(X)]}_{\text{model expected feature "count"}} = \underbrace{\mathbb{E}_{\hat{p}_n}[T_j(X)]}_{\text{empirical feature "count" "moment" constraints}}, j=1,\dots,d \}$

then

MaxENT

$$\boxed{\begin{aligned} \min_{q \in \mathbb{R}^{|X|}} \; & KL(q\|unif) \\ & q \in M \\ & q \in \Delta_{|X|} \end{aligned}}$$

$q \in M$ → $\sum_x q(x) T_j(x) = \frac{1}{n} \sum_{i=1}^n T_j(x^{(i)}) = \alpha_j$

i.e. $\langle \vec{q}, \vec{T_j} \rangle = \alpha_j$

↳ conv. opt. problem over $q \in \Delta_{|X|} \subseteq \mathbb{R}^{|X|}$

quick presentation of Lagrangian duality

convex min. problems

$$\left.\begin{aligned} \min_{x \in \mathbb{R}^d} \; & f(x) \\ & f_j(x) \le 0 \quad \forall j \in \{1,\dots,m\} \\ & g_k(x) = 0 \quad \forall k \in \{1,\dots,n\} \end{aligned}\right\} \text{"primal problem"}$$

convex opt. problem ⇐ { · f, f_j are convex fct.
· g_k affine fct.

Lagrangian fct. $\mathcal{L}(x, \lambda, \nu) \triangleq f(x) + \sum_{j=1}^m \lambda_j f_j(x) + \sum_{k=1}^n \nu_k g_k(x)$

$j=1$ ... $k=1$

"Lagrange multipliers"

magic trick (saddle pt. interpretation)

$$h(x) \triangleq \sup_{\substack{\lambda \geq 0 \\ \nu}} \mathcal{L}(x,\lambda,\nu) = \begin{cases} f(x) & \text{if } x \text{ is feasible} \\ +\infty & \text{if } x \text{ is not feasible} \end{cases}$$

$$h(x) = f(x) + \delta_M(x)$$

indicator fct. $\delta_M(x) = \begin{cases} +\infty & \text{if } x \in M \\ 0 & \text{o.w.} \end{cases}$

an equivalent problem to primal problem $\inf_x h(x)$ ← fancy non-smooth fct.

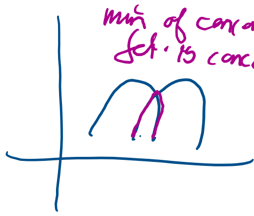
$$\inf_x \left(\sup_{\substack{\lambda \geq 0 \\ \nu}} \mathcal{L}(x,\lambda,\nu) \right)$$

duality trick is to swap inf & sup

$$\sup_{\substack{\lambda \geq 0 \\ \nu}} \underbrace{\left(\inf_x \mathcal{L}(x,\lambda,\nu) \right)}_{\triangleq\, g(\lambda,\nu)}$$ → this fct. is always concave

$g(\lambda,\nu)$ Lagrange dual fct.

min of concave fct. is concave.



Lagrange dual problem $\sup_{\substack{\lambda \geq 0 \\ \nu}} g(\lambda,\nu)$ "dual variables"

"weak duality"

in general, $\sup_{\lambda,\nu} \underbrace{\inf_x \mathcal{L}(x,\lambda,\nu)}_{g(\lambda,\nu)} \leq \inf_x \sup \mathcal{L}(x,\lambda,\nu)$

Strong duality when $\sup \inf \mathcal{L} = \inf \sup \mathcal{L}$

↳ sufficient conditions:
- primal problem is convex
- + constraint qualification condition (e.g. Slater's condition)

(can get optimal primal variables $x^*(\lambda^*,\nu^*)$ using KKT conditions)

(see ch.5 of Boyd's book)

see chapter 5 in Boyd's book for more info on duality: http://stanford.edu/~boyd/cvxbook/

13h26

$u(x) = \frac{1}{K}$

dual problem for max. entropy

MaxENT in primal form (P)

$$\min_{q \in \mathbb{R}^K} KL(q\|u) = \sum_x q(x) \log \frac{q(x)}{u(x)}$$

→ absorb this constraint in domain of def.

primal form (P)

$$\boxed{\begin{aligned} &\min_{q\in\mathbb{R}^K} \sum_x q(x)\log\frac{q(x)}{u(x)} \\ &q(x)\geq 0 \\ c \longrightarrow\ &\sum_x q(x)=1 \quad \Big\}\ \Delta_{|X|} \\ \nu \longrightarrow\ &\sum_x q(x)T_j(x)=\alpha_j \ \ \forall j \end{aligned}\ \Bigg\}\ \mathcal{M}}$$

eq. const.

absorb this constraint in domain of def.

$KL(q\|u)$ i.e.

$$KL(q\|u)=\begin{cases} +\infty & \text{if } q(x)<0 \text{ for some } x \\ KL(q\|u) & \text{o.w.}\end{cases}$$

$$\mathcal{L}(q,\nu,c)=\sum_x q(x)\log\frac{q(x)}{u(x)}+\sum_{j=1}^{d}\nu_j\Big(\alpha_j-\sum_x q(x)T_j(x)\Big)+c\Big(1-\sum_x q(x)\Big)$$

$q(x)=q_x$

$$\frac{\partial\mathcal{L}}{\partial q_x}=1+\log\frac{q(x)}{u(x)}+\sum_{j=1}^{d}\nu_jT_j(x)-c\overset{\text{want}}{=}0$$

$\sum_j \nu_j T_j(x)$

$$\Rightarrow \boxed{q^*_{\nu,c}(x)=u(x)\exp\big(\nu^\top T(x)+c-1\big)}$$

exponential family!

dual fct.:

plug back $q^*_{\nu,c}$ in $\mathcal{L}(\dots)$

$$g(\nu,c)=\mathcal{L}(q^*_{\nu,c},\nu,c)=\tilde{\mathbb{E}}_{q^*}\big[\cancel{\nu^\top T(X)}+c-1\big]+\nu^\top\alpha-\tilde{\mathbb{E}}_{q^*}\big[\cancel{\nu^\top T(X)}\big]$$

$$+c-\tilde{\mathbb{E}}_{q^*}[c]$$

shorthand $\sum_x q(x)$

$$=\nu^\top\alpha+c-\underbrace{\sum_x u(x)\exp(\nu^\top T(x))}_{\triangleq Z(\nu)}\exp(c-1)$$

$\max g(\nu,c)$ with respect to c

$$\nabla_c=0 \Rightarrow 1-\exp(c-1)Z(\nu)\overset{\text{want}}{=}0$$

$$\Rightarrow \exp(c^*-1)=\frac{1}{Z(\nu)} \quad \Rightarrow c^*-1=-\log Z(\nu)$$

plug back c^*

$$\max_c g(\nu,c)=\nu^\top\alpha+\underbrace{c^*}_{c^*-1=-\log Z(\nu)}-\cancel{Z(\nu)}\frac{1}{\cancel{Z(\nu)}}$$

dual problem $\max_\nu \tilde{g}(\nu)$ $\quad \boxed{\tilde{g}(\nu)\triangleq\nu^\top\alpha-\log Z(\nu)}$

link with MLE:

if $\alpha=\frac{1}{n}\sum_{i=1}^{n}T(x^{(i)})=\mathbb{E}_{\hat{p}_n}[T(X)]$

then $\tilde{g}(\nu)=\frac{1}{n}\sum_{i=1}^{n}\underbrace{\big[\nu^\top T(x^{(i)})-\log Z(\nu)\big]}_{\log p(x^{(i)}|\nu)+\text{cst.}}$

where $p(x|\nu)\triangleq u(x)\exp\big(\nu^\top T(x)-\log Z(\nu)\big)$

$\log p(x^{(i)}|\eta) + cst.$

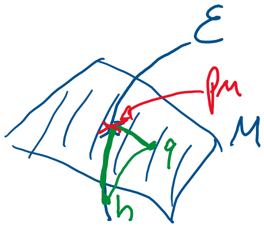
where $p(x|\eta) \triangleq u(x)\exp(\eta^T T(x) - \log Z(\eta))$

i.e. dual problem is $\max_{\eta} \tilde{g}(\eta) = \max_{\eta} \frac{1}{n} \log p(x_{1:n}|\eta) + cst.$ i.e. MLE

to summarize: $\boxed{\text{ML in exp. family with } T(x) \text{ as sufficient statistics is equivalent to max ENT with empirical moment constraints on } T(x) \text{ where } \alpha = E_{\hat{p}_n}[T(X)]}$

they are Lagrangian dual of each other!

$\boxed{\text{MLE in exp. family} \Leftrightarrow \text{moment matching in exp. family}}$



KL Pythagorean theorem

→ see lecture 16 in 2017

for any $q \in M$ and $h \in \mathcal{E}$

$KL(q||h) = KL(q||p_M) + KL(p_M||h)$

note: $\nabla_\eta \log Z(\eta) = \frac{1}{Z(\eta)} \nabla_\eta \sum_x u(x)\exp(\eta^T T(x))$

$= \sum_x T(x) \underbrace{\left(\frac{1}{Z(\eta)} u(x)\exp(\eta^T T(x))\right)}_{p(x|\eta)}$

$\nabla_\eta \log Z(\eta) = E_{p(x|\eta)}[T(x)] \triangleq \mu(\eta)$ "model moment"

$\nabla_\eta \tilde{g}(\eta) = \underbrace{\overbrace{E_{\hat{p}_n}[T(x)]}^{\alpha}}_{\hat{\mu}_n \text{ "empirical moment"}} - \mu(\eta)$

(see end of old lecture 16 2017 for "KL Pythagorean theorem" and I-projection vs. M-projection for KL + geometry)

$\nabla_\eta \tilde{g}(\eta) = 0 \Rightarrow \boxed{\mu(\eta^*) = \hat{\mu}_n}$ i.e. moment matching!