Basics

$\text{Sigmoid}(z) = \frac{1}{1+e^{-z}} = \frac{e^z}{e^z+1}$

$\text{Softmax}(z_1, \dots, z_n) = \left[\frac{e^{z_1}}{\sum e^{z_i}}, \dots, \frac{e^{z_n}}{\sum e^{z_i}}\right]$

$\ln(xy) = \ln(x) + \ln(y)$

$e^{-2\ln x} = 1/x^2$

$N(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$

Canonical: $N(x|\mu, \Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left[(x-\mu)^T \Sigma^{-1}(x-\mu)\right]$

moment: $N(x|\eta, \Omega) = \exp(a + \eta^T x - \frac{1}{2} x^T \Omega x)$

$\Omega = \Sigma^{-1}$, $\eta = \Sigma^{-1}\mu$

Linear

PSD: $x^T A x \geq 0 \quad \forall x$

gradient: $\mathbb{R}^n \to \mathbb{R}$

Jacobian: $\mathbb{R}^m \to \mathbb{R}^n$ $\quad \frac{\partial f_1}{\partial x_1} \dots \frac{\partial f_n}{\partial x_m}$

Hessian: $\mathbb{R}^m \to \mathbb{R}$ $\quad \frac{\partial f}{\partial x_1^2} \dots \frac{\partial f}{\partial x_m \partial x_1}$

Frobenius norm: $\|A\|_F = \sqrt{\sum A_{ij}^2} = \sqrt{\sum \lambda_i^2}$

spectral/L_2 norm: $\|A\|_2 = \sigma_{max}(A)$

$\lambda_{max}(A) = \sup \frac{x^T A x}{x^T x}$

diagonalization ~~[illegible]~~: $Q \Lambda Q^T$

Q ~ orthonormal eigenvecs = cols

svd: UDV^T

U cols ~ eigenvectors of XX^T

pca: $\text{Cov}(X) = UDV^T$

U cols are PCs

$\sum \lambda_i = \sum \text{Var}(K_i)$

info theory

$H(X) = -\sum_x p(x) \log p(x)$

$H(Y|X) = \sum_x p(x) \sum_y p(y|x) \log p(y|x)$

$D(p(x) \| q(x)) = \sum_x p(x) \log \frac{p(x)}{q(x)}$

$I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$

- 2 diagrams

Jensen's: $f(E(X)) \leq E(f(X))$ ~ f convex

search

$h(n)$ admissible: $h(n) \leq \text{cost}(n \to \text{goal})$

$h(n)$ consistent: $h(n) - h(n') \leq \text{cost}(n \to n')$

AC-3: repeat: apply constraints re-add neighboring arcs

stat

$\hat{\theta}_{MLE} = \arg\max_\theta p(x|\theta)$

$\hat{\theta}_{MAP} = \arg\max_\theta p(x|\theta)p(\theta)$

$\hat{\theta}_{Bayes} = \int d\theta\, \theta\, p(\theta|x)$

decisions

$EU(a|e) = \sum_{s'} P(\text{Result}(a) = s' | a, e) \cdot U(s')$

$MEU(a|e) = \max_a EU(a|e)$

$VPI(T) = E_T[MEU(a|e,T)] - MEU(a|e)$

rL

val it: $U(s) = R(s) + \gamma \max_a \sum_{s'} P(s'|s,a) U(s')$

$\pi(s) = \arg\max_\pi U(s)$

policy it: $U(s) = R(s) + \gamma \sum_{s'} P(s'|s,a) U(s')$

$\pi(s) = \arg\max_\pi U^\pi(s)$

passive

ADP: $P(s'|s,a)$, $R(s) \to$ Bellman

TD: $s \to s'$: update $U^\pi(s) = U^\pi(s) + \alpha[R(s) - U^\pi(s) + \gamma U^\pi(s')]$

active

ADP: $Q(s,a) = R(s) + \gamma \sum_{s'} P(s'|s,a) \max_{a'} Q(s',a')$

TD: $Q(s,a) = Q(s,a) + \alpha[R(s) - Q(s,a) + \gamma \max_{a'} Q(s',a')]$

SARSA: $Q(s,a) = Q(s,a) + \gamma[R(s) - Q(s,a) + \gamma Q(s',a')]$

Utility(s) = Value(s) = max_a Q(s, a)

linear reg.

sgd: $\hat{\theta} = \hat{\theta} + \alpha(y - \hat{\theta}^T x)x$

requires $0 < \alpha < \lambda_{max}[X^T X]$

Logistic reg

$P(Y=1|x) = \sigma(\omega^T x)$

Cross-entropy: $-\sum p(x) \log q(x)$

svms

soft margin: $\min_{\omega,b,e} \frac{1}{2}\|\omega\|_2^2 + C\sum_i e_i$

s.t. $y_i(\omega^T x_i - b) \geq 1 - e_i \quad \forall i$

$e_i \geq 0 \quad \forall i$

binary: $\min_{\omega,b} \frac{1}{2}\|\omega\|_2^2 + C\sum_i \max(1 - y_i(\omega^T x_i - b), 0)$

decision tree

info gain: $H(\text{parent}) - [\text{weighted ave. } H(\text{children})]$

nearest nbr

K-d tree

Locality-sensitive hashing

opt

$0 \preceq mI \preceq \nabla^2 f(x) \preceq MI$

newton: $\theta = \theta - \nabla^2 f(\theta)^{-1} \nabla f(\theta)$