Compendium

Anna Choromanska
achoroma@gmail.com
http://cims.nyu.edu/~achoroma/

Department of Electrical and Computer Engineering
New York University Tandon School of Engineering

Useful compendium

v, θ - d-dimensional vectors, H - $d \times d$ matrix

$$\|v\| = \sqrt{\sum_{i=1}^{d} v(i)^2}; \quad \|v\|^2 = v^\top v$$

$$\nabla_\theta v^\top \theta = v^\top$$

$$\nabla_\theta \theta^\top \theta = 2\theta^\top$$

$$\boxed{\nabla_\theta \theta^\top H \theta = \theta^\top (H + H^\top)}$$

$A^\top B = (B^\top A)^\top$
For invertible matrices A and B: $(AB)^{-1} = B^{-1} A^{-1}$
For orthogonal matrix A: $A^\top = A^{-1}$

Useful compendium

Let A be a square matrix. Let A and B be invertible.

$$tr(A) = \sum_{d=1}^{D} A(d,d) = tr(A^\top)$$

$$tr(AB) = tr(BA)$$

$$tr(BAB^{-1}) = tr(A)$$

$$x^\top Ax = tr(x^\top Ax) = tr(xx^\top A)$$

$$\frac{\partial tr(BA)}{\partial A} = B^\top$$

$$\frac{\partial \log(|A|)}{\partial A} = (A^{-1})^\top$$

$$(A^{-1})^{-1} = A$$

$$\det(A^{-1}) = (\det(A))^{-1} \quad \text{(or in other notation : } \quad |A^{-1}| = |A|^{-1})$$

Useful compendium

- If $p(x, y)$ is the probability density function function (pdf), i.e. density of a continuous random variable, then $\int_{x=-\infty}^{+\infty} \int_{y=-\infty}^{+\infty} p(x, y) dx dy = 1$.
- If $p(x, y)$ is the probability mass function function (pmf) (discrete random variable) then $\sum_x \sum_y p(x, y) = 1$.
- Marginal distribution $p(x)$ is $p(x) = \sum_y p(x, y)$.
- Conditional distribution is $p(x|y) = \frac{p(x,y)}{p(y)}$.
- Expectation of $f(x)$ in case of continuous random variable is $\mathbb{E}_{p(x)}[f(x)] = \int_x p(x) f(x) dx$.
- Expectation of $f(x)$ in case of discrete random variable is $\mathbb{E}_{p(x)}[f(x)] = \sum_x p(x) f(x)$.
- If a is a constant then: $\mathbb{E}[af(x)] = a\mathbb{E}[f(x)]$
- If a is a constant then: $\mathbb{E}[f(x) + a] = \mathbb{E}[f(x)] + a$
- $\mathbb{E}[\mathbb{E}[f(x)]] = \mathbb{E}[f(x)]$
- Conditional expectation $\mathbb{E}[y|x] = \int_y p(y|x) y dy$
- $\mathbb{E}[\mathbb{E}[y|x]] = \mathbb{E}[y]$

Useful compendium

- *Mean* value of x: $\mathbb{E}_{p(x)}[x]$
- *Variance* of x: $\mathrm{Var}(x) = \mathbb{E}[(x - \mathbb{E}[x])^2] = \mathbb{E}[x^2] - (\mathbb{E}[x])^2$
- *Covariance* - measure of the variability of x and y together: $\mathrm{Cov}(x, y) = \mathbb{E}[(x - \mathbb{E}[x])(y - \mathbb{E}[y])] = \mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y]$
- Sample expectation: $\mathbb{E}_{p(x)}[f(x)] = \frac{1}{N}\sum_{i=1}^{N} f(x_i)$
- Sample mean (when pdf/pmf are unknown): $\bar{x} = \mathbb{E}[x] = \frac{1}{N}\sum_{i=1}^{N} x_i$
- Sample variance: $\mathbb{E}[(x - \mathbb{E}[x])^2] = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$
- Sample covariance: $\mathbb{E}[(x - \mathbb{E}[x])(y - \mathbb{E}[y])] = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})$
- If x and y are independent:
    - $p(x, y) = p(x)p(y)$
    - $p(x|y) = p(x)$
    - $\mathbb{E}[xy] = \mathbb{E}[x]\mathbb{E}[y]$
- If x is independent of y given z, $p(x|y, z) = p(x|z)$, but $p(x|y) \neq p(x)$

Useful compendium

i.i.d. assumption:

- **independent** - probability of the data given the model (let Θ denotes model parameters) multiplies:

$$p(x_1, x_2, \ldots, x_N|\Theta) = \prod_{i=1}^{N} p_i(x_i|\Theta)$$

- **identically distributed** - marginal probabilities are the same for each data point

$$p(x_1, x_2, \ldots, x_N|\Theta) = \prod_{i=1}^{N} p_i(x_i|\Theta) = \prod_{x=1}^{N} p(x_i|\Theta)$$

- One can learn joint distribution by maximum log-likelihood:

$$\Theta^* = \arg\max_{\Theta} \log \prod_{i=1}^{N} p(x_i|\Theta) = \arg\max_{\Theta} \sum_{i=1}^{N} \log p(x_i|\Theta)$$

- One can learn conditional distribution by maximum conditional log-likelihood:

$$\Theta^* = \arg\max_{\Theta} \log \prod_{i=1}^{N} p(y_i|x_i, \Theta) = \arg\max_{\Theta} \sum_{i=1}^{N} \log p(y_i|x_i, \Theta)$$