Standard formula sheet

1. Distributions:

Normal $f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Binomial - $B(n,p)$ $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$

Poisson $P(\mathrm{X} = k) = e^{-\lambda} \frac{\lambda^k}{k!}$

Geometric $P(X = k) = (1-p)^{k-1} p$

2. Decision Trees:

Gini $Gini(S) = 1 - \sum_{i=1}^{c} \left(\frac{|S_i|}{|S|}\right)^2$

Entropy $Entropy(S) = -\sum_{i=1}^{c} \frac{|S_i|}{|S|} \log \frac{|S_i|}{|S|}$

3. Gradient descent and update steps:

Linear regression $\theta_j \coloneqq \theta_j - \alpha \frac{1}{m} \sum_{d \in D} \left(h_\theta\left(x^{(d)}\right) - y^{(d)}\right) \cdot x_j^{(d)}$

Perceptron $w_j \coloneqq w_j - \eta \sum_{d \in D} (o^{(d)} - t^{(d)}) x_j^{(d)}$

Dual perceptron If $o^{(d)} \cdot t^{(d)} < 0$ then:

$\alpha_j = \alpha_j + \eta$

4. Logistic regression:

$P(h(x) = 1) = \frac{1}{1 + e^{-w^T x}}$

5. SVM:

Primal objective function $\frac{1}{2}\|w\|^2 + \gamma \sum_d \xi_d - \sum_d \alpha_d (t_d(w^T x_d + w_0) - 1 + \xi_d) - \sum_d \mu_d \xi_d$

s.t $\alpha_d \geq 0 \quad \mu_d \geq 0$

Dual objective function $\sum_d \alpha_d - 1/2 \sum_d \sum_e \alpha_d \alpha_e t_d t_e x_d^T x_e$

s.t $\sum_d \alpha_d t_d = 0,\ 0 \leq \alpha_d \leq \gamma$

6. EM (for Bernoulli distributions):

$New\ w_{A_j} = \frac{1}{N} \sum_{i=1}^{N} r(x_i, A_j)$

$p_{A_j} = \frac{1}{(New\ w_{A_j})N} \sum_{i=1}^{N} r\left(x_i, A_j\right) v(i)$

7. Linear Regression (closed form): $\theta^* = \underset{\theta}{\operatorname{argmin}} \|y - X \cdot \theta\|_2^2 = (X^T X)^{-1} X^T y$