Formula Sheet

Integration

$$\int_a^b x^d dx = \frac{x^{d+1}}{d+1}\Big|_a^b = \frac{b^{d+1} - a^{d+1}}{d+1} \quad \text{for } d \neq -1$$

Derivatives and Gradient

$$f(x) = x^a, \qquad f'(x) = \frac{df}{dx}(x) = ax^{a-1}$$

$$f(x) = \exp(x), \qquad f'(x) = \frac{df}{dx}(x) = \exp(x)$$

$$f(x) = \ln(x), \qquad f'(x) = \frac{df}{dx}(x) = \frac{1}{x}$$

$$f(x) = g(h(x)), \quad u = h(x) \qquad f'(x) = \frac{df}{dx}(x) = \frac{dg}{du}\frac{du}{dx}(x) = g'(u)h'(x) \quad \triangleright \text{Chain rule}$$

$$f(x) = g(x)h(x), \qquad f'(x) = \frac{df}{dx}(x) = g'(x)h(x) + g(x)h'(x) \quad \triangleright \text{Product rule}$$

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}(\mathbf{x}), \frac{\partial f}{\partial x_2}(\mathbf{x}), \dots, \frac{\partial f}{\partial x_d}(\mathbf{x})\right)^\top \quad \text{for } \mathbf{x} \in \mathbb{R}^d$$

Probability

Univariate: $\mathbb{P}(X \in \mathcal{E}) \overset{\text{def}}{=} \begin{cases} \sum_{x \in \mathcal{E}} p(x) & \text{if } X \text{ is discrete} \\ \int_{\mathcal{E}} p(x)dx & \text{if } X \text{ is continuous} \end{cases}$

Multivariate: $\mathbb{P}(X \in \mathcal{E}_X, Y \in \mathcal{E}_Y) \overset{\text{def}}{=} \begin{cases} \sum_{x \in \mathcal{E}_x} \sum_{y \in \mathcal{E}_y} p(x,y) & \text{if } X, Y \text{ are discrete} \\ \int_{\mathcal{E}_x} \int_{\mathcal{E}_y} p(x,y)dydx & \text{if } X, Y \text{ are continuous} \\ \int_{\mathcal{E}_x} \sum_{y \in \mathcal{E}_y} p_{Y|X}(y|x)p_X(x)dx & \text{if } X \text{ is continuous, } Y \text{ is discrete} \\ \sum_{x \in \mathcal{E}_x} \int_{\mathcal{E}_y} p_{Y|X}(y|x)p_X(x)dy & \text{if } X \text{ is discrete, } Y \text{ is continuous} \end{cases}$

Marginal pmf: $p_X(x) \overset{\text{def}}{=} \begin{cases} \sum_{y \in \mathcal{Y}} p(x,y) & \text{if } Y \text{ is discrete} \\ \int_{\mathcal{Y}} p(x,y)dy & \text{if } Y \text{ is continuous} \end{cases}$

Marginal: $\mathbb{P}_X(X \in \mathcal{E}_X) \overset{\text{def}}{=} \begin{cases} \sum_{x \in \mathcal{E}_X} p_X(x) & \text{if } X \text{ is discrete} \\ \int_{\mathcal{E}_X} p_X(x)dx & \text{if } X \text{ is continuous} \end{cases}$

Conditional pmf: $p_{Y|X}(y|x) \overset{\text{def}}{=} \frac{p(x,y)}{p_X(x)}$ such that $p_X(x) > 0$

Conditional: $\mathbb{P}_{Y|X}(Y \in \mathcal{E}_Y | X = x) \overset{\text{def}}{=} \begin{cases} \sum_{y \in \mathcal{E}_Y} p_{Y|X}(y|x) & \text{if } Y \text{ is discrete} \\ \int_{\mathcal{E}_Y} p_{Y|X}(y|x)dy & \text{if } Y \text{ is continuous} \end{cases}$

Product Rule: $p(x,y) = p_{Y|X}(y|x)p_X(x)$

Bayes' Rule: $p_{X|Y}(x|y) = \frac{p_{Y|X}(y|x)p_X(x)}{p_Y(y)}$

Independence: $p(x_1, \dots, x_n) = p_{X_1}(x_1) \cdots p_{X_n}(x_n)$

Distribution	Parameters	pmf or pdf	Expectation and Variance
Bernoulli	$\alpha \in [0,1]$	$p(x) = \alpha^x(1-\alpha)^{1-x}$, $x \in \{0,1\}$	$\mathbb{E}[X] = \alpha$, $\text{Var}[X] = \alpha(1-\alpha)$
Discrete Uniform	$n \in \mathbb{N}$	$p(x) = \frac{1}{n}$, $x \in \{1, \dots, n\}$	$\mathbb{E}[X] = \frac{n+1}{2}$, $\text{Var}[X] = \frac{n^2-1}{12}$
Continuous Uniform	$a, b \in \mathbb{R}, a < b$	$p(x) = \frac{1}{b-a}$, $x \in [a,b]$	$\mathbb{E}[X] = \frac{a+b}{2}$, $\text{Var}[X] = \frac{(b-a)^2}{12}$
Normal	$\mu \in \mathbb{R}, \sigma^2 > 0$	$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$, $x \in \mathbb{R}$	$\mathbb{E}[X] = \mu$, $\text{Var}[X] = \sigma^2$
Laplace	$\mu \in \mathbb{R}, b > 0$	$p(x) = \frac{1}{2b} \exp\left(-\frac{\lvert x-\mu \rvert}{b}\right)$, $x \in \mathbb{R}$	$\mathbb{E}[X] = \mu$, $\text{Var}[X] = 2b^2$
Categorical	$\alpha_0, \dots, \alpha_{K-1} \in [0,1]$	$p(x) = \alpha_x$, $x \in \{0, \dots, K-1\}$	Not useful

Expectation and Variance

Univariate: $\mathbb{E}[X] \overset{\text{def}}{=} \begin{cases} \sum_{x\in\mathcal{X}} xp(x) & \text{if } X \text{ is discrete} \\ \int_{\mathcal{X}} xp(x)dx & \text{if } X \text{ is continuous} \end{cases}$

Function: $\mathbb{E}[f(X)] \overset{\text{def}}{=} \begin{cases} \sum_{x\in\mathcal{X}} f(x)p(x) & \text{if } X \text{ is discrete} \\ \int_{\mathcal{X}} f(x)p(x)dx & \text{if } X \text{ is continuous} \end{cases}$

Variance: $\mathrm{Var}[X] \overset{\text{def}}{=} \mathbb{E}\left[(X-\mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2$

Multivariate: $\mathbb{E}[f(X,Y)] \overset{\text{def}}{=} \begin{cases} \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} f(x,y)p(x,y) & \text{if } X \text{ and } Y \text{ are discrete} \\ \int_{\mathcal{X}}\int_{\mathcal{Y}} f(x,y)p(x,y)\,dy\,dx & \text{if } X \text{ and } Y \text{ are continuous} \\ \int_{\mathcal{X}}\sum_{y\in\mathcal{Y}} f(x,y)p_{Y|X}(y|x)p_X(x)\,dx & \text{if } X \text{ is continuous, } Y \text{ is discrete} \\ \sum_{x\in\mathcal{X}}\int_{\mathcal{Y}} f(x,y)p_{Y|X}(y|x)p_X(x)\,dy & \text{if } X \text{ is discrete, } Y \text{ is continuous} \end{cases}$

Conditional: $\mathbb{E}[f(Y)|X=x] \overset{\text{def}}{=} \begin{cases} \sum_{y\in\mathcal{Y}} f(y)p_{Y|X}(y|x) & \text{if } Y \text{ is discrete} \\ \int_{\mathcal{Y}} f(y)p_{Y|X}(y|x)\,dy & \text{if } Y \text{ is continuous} \end{cases}$

Expectation and Variance Properties

1. $\mathbb{E}[cX] = c\mathbb{E}[X]$
2. $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
3. $\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$
4. $\mathrm{Var}[c] = 0$
5. $\mathrm{Var}[cX] = c^2\mathrm{Var}[X]$.

If X and Y are independent:

6. $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$
7. $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$.

Estimation

Sample Mean: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ **If X_i are i.i.d.:** $\mathbb{E}[\bar{X}] = \mathbb{E}[X_1]$, $\mathrm{Var}\left[\bar{X}\right] = \frac{\mathrm{Var}[X_1]}{n}$

Estimate of Expected Loss: $\hat{L}(f) = \frac{1}{n}\sum_{i=1}^{n} \ell(f(\boldsymbol{X}_i), Y_i)$

Optimization

2nd Order GD: $w^{(t+1)} = w^{(t)} - \frac{g'(w^{(t)})}{g''(w^{(t)})}$ **1st Order GD:** $w^{(t+1)} = w^{(t)} - \eta^{(t)}g'(w^{(t)})$ **Poly Dim:** $\bar{p} = \binom{d+p}{p}$

Evaluation

EE and AE: $$\mathbb{E}[L(\hat{f}_D)] = \underbrace{\mathbb{E}[L(\hat{f}_D)] - L(f^*)}_{\text{Estimation Error (EE)}} + \underbrace{L(f^*) - L(f_{\text{Bayes}})}_{\text{Approximation Error (AE)}} + \underbrace{L(f_{\text{Bayes}})}_{\text{Irreducible Error (IE)}}$$

Bias and Var: $$\mathbb{E}[L(\hat{f}_D)] = \mathbb{E}\left[\underbrace{\mathbb{E}[(\hat{f}_D(\boldsymbol{X}) - \bar{f}(\boldsymbol{X}))^2|\boldsymbol{X}]}_{\text{Variance}}\right] + \mathbb{E}\left[\underbrace{(\bar{f}(\boldsymbol{X}) - f_{\text{Bayes}}(\boldsymbol{X}))^2}_{\text{Bias}}\right] + \underbrace{L(f_{\text{Bayes}})}_{\text{Irreducible Error}}$$

Regularization: $$\hat{L}_\lambda(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i^\top\mathbf{w} - y_i)^2 + \frac{\lambda}{n}\sum_{j=1}^{d} w_j^2$$

MLE and MAP

MLE: $\arg\max_{w\in\mathcal{W}} p(\mathcal{D}|w)$ **MAP:** $\arg\max_{w\in\mathcal{W}} p(w|\mathcal{D}) = \arg\max_{w\in\mathcal{W}} p(\mathcal{D}|w)p(w)$

Squared loss: $f_{\text{Bayes}}(\mathbf{x}) = \mathbb{E}[Y|\boldsymbol{X}=\mathbf{x}] = \int_{\mathcal{Y}} y\cdot p_{Y|\mathbf{X}}(y|\mathbf{x})\,dy$, **Zero-One loss:** $f_{\text{Bayes}}(\mathbf{x}) = \arg\max_{y\in\mathcal{Y}} p(y|\mathbf{x})$.

Classification

Logistic: $\sigma(z) = \frac{1}{1+\exp(-z)}$ **Softmax:** $\sigma(\mathbf{z}) = (\sigma_0(\mathbf{z}),\ldots,\sigma_{K-1}(\mathbf{z}))^\top$ where $\sigma_y(\mathbf{z}) = \frac{\exp(z_y)}{\sum_{q=0}^{K-1}\exp(z_q)}$

Neural Networks

Weights: $\mathbf{w}_1^{(b)},\ldots,\mathbf{w}_{d^{(b)}}^{(b)} \in \mathbb{R}^{d^{(b-1)}+1}$ **Pre-acts:** $\mathbf{z}^{(b)} = \left(z_1^{(b)},\ldots,z_{d^{(b)}}^{(b)}\right) \in \mathbb{R}^{d^{(b)}}$, $z_j^{(b)} = \left(\mathbf{a}^{(b-1)}\right)^\top \mathbf{w}_j^{(b)}$ **Act func:** $h^{(b)}:\mathbb{R}\to\mathbb{R}$

Activations: $\mathbf{a}^{(0)} = \mathbf{x} \in \mathbb{R}^{d+1} = \mathbb{R}^{d^{(0)}+1}$, $\mathbf{a}^{(b)} = \left(a_0^{(b)} = 1, a_1^{(b)},\ldots,a_{d^{(b)}}^{(b)}\right) \in \mathbb{R}^{d^{(b)}+1}$ except $b=B$, $\mathbf{a}^{(B)} = \left(a_1^{(B)},\ldots,a_{d^{(B)}}^{(B)}\right) \in \mathbb{R}^{d^{(B)}}$

where $a_j^{(b)} = h^{(b)}(z_j^{(b)})$ **Number of Layers:** B, **Number of Non-Bias Neurons:** $d^{(b)}$