Question:

$$f(\vec{x}, \vec{w}) = \overbrace{\vec{x}^T\vec{w}}^{\text{linear function}} \quad \text{where } \vec{x}, \vec{w} \in \mathbb{R}^d$$

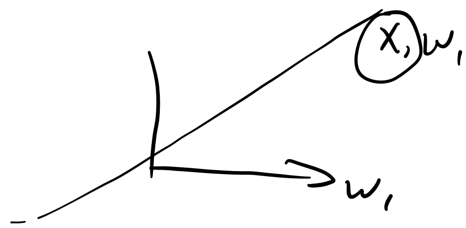
What is $(\nabla_{\vec{w}} f)(\vec{w}) = \left(\frac{\partial f}{\partial w_1}(w_1), \ldots, \frac{\partial f}{\partial w_d}(w_d)\right)^T$?

gradient

Solution:

$\vec{x} = (x_1, \ldots, x_d)^T, \ \vec{w} = (w_1, \ldots, w_d)^T$

$$\begin{aligned}
f(\vec{x}, \vec{w}) &= f(x_1, \ldots, x_d, w_1, \ldots, w_d) \\
&= \vec{x}^T\vec{w} \\
&= (x_1, \ldots, x_d)\begin{pmatrix} w_1 \\ \vdots \\ w_d \end{pmatrix} = (x_1, \ldots, x_d)(w_1, \ldots, w_d)^T \\
&= x_1 w_1 + x_2 w_2 + \ldots + x_d w_d \\
&= \sum_{i=1}^{d} x_i w_i
\end{aligned}$$



$$\nabla_{\vec{w}} f(\vec{w}) = \left(\frac{\partial f}{\partial w_1}(w_1) = x_1, \ \frac{\partial f}{\partial w_2} = x_2, \ \ldots, \ \frac{\partial f}{\partial w_d} = x_d\right)^T$$

$$g(\vec{x}, \vec{w}) = \sum_{i=1}^{d} x_i w_i^2 = x_1 w_1^2 + \ldots + x_d w_d^2$$

$$\begin{aligned}
\nabla_{\vec{w}} g(\vec{w}) &= \left(\frac{\partial g}{\partial w_1}(w_1) = 2 w_1 x_1, \ \ldots, \ \frac{\partial g}{\partial w_d}(w_d) = 2 w_d x_d\right)^T \\
&= \nabla_{\vec{w}} g(w_1, \ldots, w_d)
\end{aligned}$$

Question:

$$\ell(\hat{y}, y) = (\hat{y} - y)^2$$

$$f(\vec{x}, \vec{w}) = \vec{x}^T \vec{w} \quad \text{where } \vec{x}, \vec{w} \in \mathbb{R}^d$$

$$\ell(\hat{y}, y) = \exp(\hat{y}) - y \quad \text{where } \hat{y}, y \in \mathbb{R}$$

$$h(\vec{x}, \vec{w}, y) = \ell(f(\vec{x}, \vec{w}), y) = \exp(\vec{x}^T \vec{w}) - y$$

What is $\nabla_{\vec{w}} h(\vec{w}) = \left(\frac{\partial h}{\partial w_1}(w_1), \ldots, \frac{\partial h}{\partial w_d}(w_d) \right)^T$?

Solution:

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

$$\exp(z) = e^z$$

$$\exp(\vec{x}^T \vec{w}) = e^{\vec{x}^T \vec{w}}$$

$$\begin{aligned}
h(\vec{x}, \vec{w}, y) &= h(x_1, \ldots, x_d, w_1, \ldots, w_d, y) \\
&= \exp(\vec{x}^T \vec{w}) - y \\
&= \exp(x_1 w_1 + \ldots + x_d w_d) - y \\
&= \exp\left(\sum_{i=1}^{d} x_i w_i\right) - y
\end{aligned}$$

$$f = f(\vec{x}, \vec{w}) \in \mathbb{R}$$

$$\in \{f \mid f : \mathcal{X} \to \mathcal{Y}\}$$

$$\hat{y} = f(\vec{x}, \vec{w})$$

$$h(\hat{y}, y) = \exp(\hat{y}) - y$$

$$\frac{\partial h}{\partial \hat{y}}(\hat{y}) = \exp(\hat{y})$$

$$\frac{\partial \hat{y}}{\partial w_1} = \frac{\partial f}{\partial w_1} = x_1$$

$$\left(\frac{\partial h}{\partial w_1}\right)(w_1) = \left(\frac{\partial h}{\partial \hat{y}} \; \frac{\partial \hat{y}}{\partial w_1}\right)(w_1)$$

$$= \exp(f(\vec{x},\vec{w}) \cdot x_1$$

$$= \exp\left(\sum_{i=1}^{d} x_i w_i\right) \cdot x_1$$

$$\nabla_{\vec{w}} h(\vec{w}) = \left(\exp\left(\sum_{i=1}^{d} x_i w_i\right) x_1 , \ldots , \exp\left(\sum_{i=1}^{d} x_i w_i\right) x_d \right)^T$$

What happens when $Z=(X,Y)$ with Y discrete and X continuous?

$\mathbb{P}_Z \quad p$

$p: \mathcal{X} \times \mathcal{Y} \Rightarrow ?$ pmf or pdf? Ans: neither

Instead we will write $p(x,y)$ in terms of a marginal pdf for X and a conditional pmf for $Y|X$

$$p(x,y) = p_X(x)\, p_{Y|X}(y|x)$$ product rule

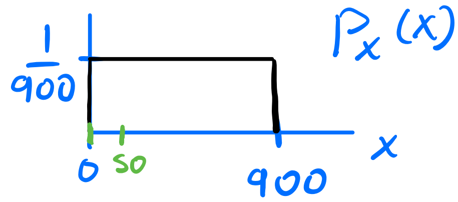
where $p_{Y|X=x}: \mathcal{Y} \Rightarrow [0,1]$ is a pmf

$p_X(x): \mathcal{X} \Rightarrow [0,\infty)$ is a pdf

Ex: $X \in \mathcal{X} = [0,900]$, $Y \in \mathcal{Y} \in \{0,1\}$ Barolo

pdf: $\mathbb{P}_X = \text{Uniform}(0,900)$

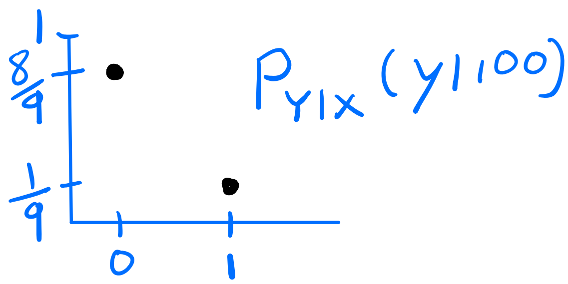
$p_X(x) = \frac{1}{900}$



$\mathbb{P}_{Y|X=x} = \text{Bernoilli}\left(\frac{x}{900}\right)$

pmf: $p_{Y|X}(y|x) = \begin{cases} \frac{x}{900} & \text{if } y=1 \\ 1-\frac{x}{900} & \text{if } y=0 \end{cases}$

Defn of Bernoilli$\left(\frac{x}{900}\right)$



$p_{Y|X}(1|900) = 1$

what is the probability a wine $Z=(X,Y)$ has 0 to 50 mg of a chemical and is a Barolo wine

What is

$$Z \in \mathcal{Z} = \mathcal{X} \times \mathcal{Y} = \{(x,y) \mid x \in \mathcal{X}, y \in \mathcal{Y}\}$$

$$\mathbb{P}_Z(Z \in \mathcal{E}) \qquad \mathcal{E} = \{(x,y) \mid x \in [0,50], y \in \{1\}\}$$

$$= \underbrace{[0,50]}_{=\mathcal{E}_x} \times \underbrace{\{1\}}_{=\mathcal{E}_y}$$

$$\mathbb{P}_Z(X \underset{\in \mathcal{E}_x}{\in} [0,50], Y \underset{\in \mathcal{E}_y}{=} 1) = \int_{[0,50]} \left(\sum_{y \in \{1\}} p(x,y) \right) dx$$

$$= \int_0^{50} \left(\sum_{y \in \{1\}} p_{Y|X}(y|x)\, p_X(x) \right) dx$$

$$= \int_0^{50} \left(\frac{x}{900} \; \frac{1}{900} \right) dx$$

$$= \frac{1}{900^2} \int_0^{50} x \, dx$$

$$= \frac{1}{900^2} \left. \frac{x^2}{2} \right|_0^{50}$$

$$= \frac{1}{900^2} \left(\frac{50^2}{2} - \frac{0^2}{2} \right)$$

$$= \frac{1}{810000} \cdot \frac{2500}{2}$$

Multivariate Expected Value:

$Z = (X, Y) \in \mathcal{X} \times \mathcal{Y} = \mathcal{Z}$ is a r.v.

$f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$

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

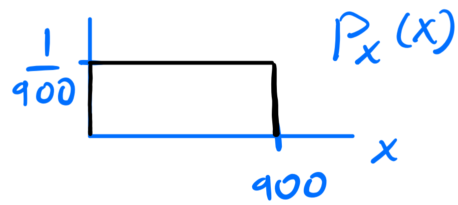
you can always use:

$$p(x,y) = p(y|x)p(x) = p(x|y)\,p(x)$$

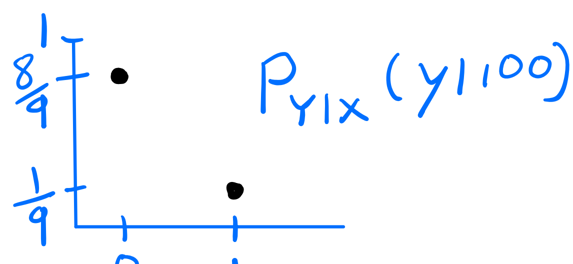
Ex: $X \in \mathcal{X} = [0, 900]$, $Y \in \mathcal{Y} \in \{0, 1\}$ ← Barolo

pdf: $p_X = \text{Uniform}(0, 900)$

$= \frac{1}{900}$



$P_{Y|X=x} = \text{Bernoilli}\left(\frac{x}{900}\right)$



pmf: $P_{Y|X}(y|x) = \begin{cases} \frac{x}{900} & \text{if } y = 1 \end{cases}$

$\left| 1-\frac{x}{900} \right.$ if $y=0$

Defn of Bernoilli$\left(\frac{x}{900}\right)$

$$\ell(X,Y) = \left(\frac{X}{900} - Y\right)^2$$

what is

$$\mathbb{E}[\ell(X,Y)] = \int_{\mathcal{X}} \left(\sum_{y \in \mathcal{Y}} \ell(x,y) p(y|x) \right) p(x)\, dx$$

$$= \int_0^{900} \left(\sum_{y \in \{0,1\}} \left(\frac{x}{900} - y\right)^2 p(y|x) \right) p(x)\, dx$$

$$= \int_0^{900} \left(\left(\frac{x}{900} - 0\right)^2 \left(1 - \frac{x}{900}\right) + \left(\frac{x}{900} - 1\right)^2 \left(\frac{x}{900}\right) \right) \frac{1}{900}\, dx$$

$$= \frac{1}{900} \int_0^{900} \frac{x}{900} \left(1 - \frac{x}{900}\right) dx$$

$$= \frac{1}{900} \int_0^{900} \left(\frac{x}{900} - \frac{x^2}{900^2} \right) dx$$

$$= \frac{1}{900} \int_0^{900} \frac{x}{900}\, dx - \frac{1}{900} \int_0^{900} \frac{x^2}{900^2}\, dx$$

$$= \frac{1}{900^2} \left. \frac{x^2}{2} \right|_0^{900} - \frac{1}{900^3} \left. \frac{x^3}{3} \right|_0^{900}$$

$$= \frac{1}{6}$$

Expected value of functions of r.v.:

$X \in \mathcal{X}$ is a r.v. with pmf or pdf p

The function $f: \mathcal{X} \to \mathcal{Y}$ must have $\mathcal{Y} = \mathbb{R}$

Law of the unconscious statistician (LOTUS)

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

Ex: (X is the payout from a slot machine)

$X \in [-10, 10]$ with $p(x) = \frac{1}{20}$, $\mathbb{P} = \text{Uniform}(-10, 10)$

$Y = f(X) = X^2 \in [0, 100] = \mathcal{Y}$

$p_Y(y) = \frac{1}{20\sqrt{y}}$ much more complicated

$$\mathbb{E}[f(X)] = \int_{\mathcal{X}} f(x)p(x)\,dx = \int_{-10}^{10} x^2 \frac{1}{20}\,dx$$

$$= \frac{X^3}{3} \cdot \frac{1}{20} \Big|_{-10}^{10}$$

$$= \left(\frac{1000}{3} - \left(\frac{-1000}{3}\right)\right) \cdot \frac{1}{20}$$

$$= \frac{2000}{60} = 33.333$$

It turns out

$$\mathbb{E}[f(x)] = \mathbb{E}[Y] = \int_y Y \, p_Y(y) \, dy$$

exercise $\rightarrow$ $= 33.333$

Usually we don't know $p_Y = p_{f(x)}$

So we work with p