数式を中心に復習する

<div style="page-break-after: always;"></div>

固有方程式

$$\det(\lambda I - A) = 0$$

各要素の意味

- $\det$ は行列式(determinant) を意味する。
- λ は固有値
- I は
- A は

<div style="page-break-after: always;"></div>

特異値分解

$$A = U\Sigma V^T$$

各要素の意味

- A は
- U は
- Σ は
- V^T は
-

<div style="page-break-after: always;"></div>

ベルヌーイ分布

$$Ef(x;p) = p^x(1-p)^{1-x}$$

各要素の意味

- x　は
- p　は
- $f(x;p)$　は
- $p^x(1-p)^{1-x}$　は

<div style="page-break-after: always;"></div>

ベルヌーイ分布の期待値

$$\mathbb{E}[X] = \sum_{x=0}^{1} x p^x (1-p)^{1-x}$$

$$= p$$

各要素の意味

<div style="page-break-after: always;"></div>

ベルヌーイ分布の分散

$$Var[X] = \mathbb{E}[(X - \mathbb{E}[X])^2)]$$

$$\begin{aligned}
Var[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\
&= \sum_{x=0}^{1} x^2 p^x (1-p)^{1-x} - p^2 \\
&= p - p^2 \\
&= p(1-p)
\end{aligned}$$

各要素の意味

<div style="page-break-after: always;"></div>

マルチヌーイ分布（カテゴリ分布）

$$f(x;p) = \prod_{j=1}^{k} p_j^{x_j} \quad (\text{ただし、} \sum_{j=1}^{k} p_j = 1\text{、}0 \leq p_j \leq 1\text{、}j = 1, \ldots.)$$

各要素の意味

<div style="page-break-after: always;"></div>

マルチヌーイ分布（カテゴリ分布）の負の対数尤度

$$
\begin{aligned}
-\log L_D(p) &= -\log \prod_{i=1}^{n} f(x_i; p) \\
&= -\sum_{i=1}^{n} \log \prod_{j=1}^{k} p_j^{x_{ij}} \\
&= -\sum_{i=1}^{n} \sum_{j=1}^{k} \log p_j^{x_{ij}} \\
&= -\sum_{i=1}^{n} \sum_{j=1}^{k} x_{ij} \log p_j
\end{aligned}
$$

正規分布

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

各要素の意味

<div style="page-break-after: always;"></div>

正規分布の負の対数尤度

$$
\begin{aligned}
L(\mu) &= \prod_{i=1}^{n} f(x_i; \mu) \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_i - \mu)^2\right) \\
-\log L(\mu) &= -\log\left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_i - \mu)^2\right)\right) \\
&= -\sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_i - \mu)^2\right)\right) \\
&= -\sum_{i=1}^{n} \left(\log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{1}{2}(x_i - \mu)^2\right) \\
&= -n \log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2
\end{aligned}
$$

正規分布の最尤推定

$$
\begin{aligned}
\frac{d}{d\mu} g(\mu) &= \frac{1}{2} \sum_{i=1}^{n} \frac{d}{d\mu} (x_i - \mu)^2 \\
&= \frac{1}{2} \sum_{i=1}^{n} (-2(x_i - \mu)) \\
&= \sum_{i=1}^{n} \mu - \sum_{i=1}^{n} x_i \\
&= n\mu - \sum_{i=1}^{n} x_i
\end{aligned}
$$

$$
\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i
$$

エントロピー

$$H(X) = -\sum_{x} p(x) \log_2 p(x)$$

各要素の意味

<div style="page-break-after: always;"></div>

交差エントロピー(クロスエントロピー)の定義

$$H(p,q) = -\sum_{x} p(x)\log_2 q(x)$$

各要素の意味

- $p(x)$　真の(正解の)確率分布
- $q(x)$　推定したモデルの確率分布

<div style="page-break-after: always;"></div>

二値交差エントロピー(バイナリクロスエントロピー)　※1/8追加※

$$D_{BC} = -P(x=0)\log Q(x=0) - (1 - P(x=0))\log(1 - Q(x-0))$$

各要素の意味

- $p(x)$　真の(正解の)確率分布
- $q(x)$　推定したモデルの確率分布

<div style="page-break-after: always;"></div>

KLダイバージェンス

$$D(p||q) = \sum_x p(x) \log_2 \frac{p(x)}{q(x)}$$

各要素の意味

<div style="page-break-after: always;"></div>

JSダイバージェンス

$$D_{JS}(p||q) = \frac{1}{2}\left(\sum_x p(x)\log_2\frac{p(x)}{r(x)} + \sum_x q(x)\log_2\frac{q(x)}{r(x)}\right)$$

$$r(x) = \frac{p(x)+q(x)}{2}$$

各要素の意味

<div style="page-break-after: always;"></div>

ベイズの定理

$$p(C|x) = \frac{p(x|C)p(C)}{p(x)}$$

各要素の意味

<div style="page-break-after: always;"></div>

バイアス・バリアンス・ノイズ

$$\mathbb{E}(L) = \int \{y(x) - h(x)\}^2 p(x)dx + \iint \{h(x) - t\}^2 p(x,t)dxdt$$

$$\int \{\mathbb{E}_D[y(x;D)] - h(x)\}^2 p(x)dx$$

$$\int \mathbb{E}_D[\{y(x;D) - \mathbb{E}_D[y(x;D)]\}^2 p(x)dx$$

$$\iint \{h(x) - t\}^2 p(x,t)dxdt$$

各要素の意味

<div style="page-break-after: always;"></div>

シグモイド関数

$$f(x) = \frac{1}{1 + \exp(-x)}$$

各要素の意味

- $\exp(-x)$ は e^{-x} の意、x=0の時1となるため、$f(x) = \frac{1}{2}$ となる
-

<div style="page-break-after: always;"></div>

ReLU

$$f(x) = \max(0, x)$$

各要素の意味

<div style="page-break-after: always;"></div>

オッズ

$$\frac{p(y=1|x)}{p(y=0|x)} = \frac{\hat{y}}{1-\hat{y}}$$

$$\begin{aligned}
\frac{\hat{y}}{1-\hat{y}} &= \frac{\frac{1}{1+\exp(-w^Tx-b)}}{1-\frac{1}{1+\exp(-w^Tx-b)}} \\
&= \frac{1}{(1+\exp(-w^Tx-b))-1} \\
&= \frac{1}{\exp(-w^Tx-b)} \\
&= \exp(w^Tx-b)
\end{aligned}$$

ガウスカーネル

$$k(x, x') = \exp\left(-\frac{||x - x'||^2}{\beta}\right)$$

各要素の意味

<div style="page-break-after: always;"></div>

正則化

$$E + \lambda_2 ||w||_2^2$$
$$E + \lambda_1 ||w||_1$$
$$E + \lambda_1 ||w||_1 + \lambda_2 ||w||_2^2$$

各要素の意味

<div style="page-break-after: always;"></div>

ソフトマックス

$$\mathrm{softmax}(z)_i = \frac{\exp(z_i)}{\Sigma_j \exp(z_j)}$$

Numpy表記： `np.exp(z) / np.sum(np.exp(z))`

各要素の意味

<div style="page-break-after: always;"></div>

二乗和誤差

$$\frac{1}{2}\sum_{k=1}^{K} = (y_k - t_k)^2$$

各要素の意味

<div style="page-break-after: always;"></div>

生成モデル

$$p(y|x)p(x) = \frac{p(x,y)}{p(x)} \cdot p(x) = p(x,y)$$

各要素の意味

<div style="page-break-after: always;"></div>

ベルマン方程式

$$V^\pi(s) = \mathbb{E}[G_t|S_t = s]$$

$$= \mathbb{E}[R_{t+1} + \gamma G_{t+1}|S_t = s]$$

$$= \sum_a \pi(a|s) \sum_{s',r} P(s',r|s,a)[r + \gamma \mathbb{E}_\pi[G_{t+1}|S_{t+1} = s']]$$

$$= \sum_a \pi(a|s) \sum_{s',r} P(s',r|s,a)[r + \gamma V^\pi(s')]$$

$$\begin{aligned} Q^\pi(s,a) &= \mathbb{E}[R_{t+1} + \gamma V^\pi(S_{t+1})|S_t = s, A_t = a] \\ &= \sum_{s',r} P(s',r|s,a)[r + \gamma V^\pi(s')] \end{aligned}$$

各要素の意味

<div style="page-break-after: always;"></div>

SARSA

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha[R_{t+1} + \gamma Q(S_{t+1}, A_{t+1}) - Q(S_t, A_t)]$$

各要素の意味

<div style="page-break-after: always;"></div>

Q学習

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha \left[R_{t+1} + \gamma \max_{a'} Q(S_{t+1}, a') - Q(S_t, A_t) \right]$$

各要素の意味

<div style="page-break-after: always;"></div>

方策勾配定理

$$\nabla_\theta J(\theta) = \sum d^{\pi_\theta}(s) \sum \nabla_\theta \pi_\theta(a|s,\theta) Q^{\pi_\theta}(s,a)$$

$$\nabla_\theta \log \pi_\theta(a|s) = \frac{\partial \pi_\theta(a|s)}{\partial \theta} \frac{1}{\pi_\theta(a|s)}$$

$$d^{\pi_\theta}(s) = \sum_{k=0}^{\infty} \gamma^k P^{\pi_\theta}(s_k = s|s_0)$$

$$\begin{aligned}
\nabla_\theta J(\theta) &= \sum d^{\pi_\theta}(s) \sum_a (\nabla_\theta \pi_\theta(a|s,\theta)) Q^{\pi_\theta}(s,a) \\
&= \sum d^{\pi_\theta}(s) \sum_a \pi_\theta(a|s,\theta)(\nabla_\theta \log \pi_\theta(a|s,\theta)) Q^{\pi_\theta}(s,a) \\
&= \mathbb{E}_{\pi_\theta}[\nabla_\theta \log \pi_\theta(a|s,\theta) Q^{\pi_\theta}(s,a)]
\end{aligned}$$

各要素の意味

<div style="page-break-after: always;"></div>

モンテカルロ近似

$$\nabla_\theta J(\theta) = \mathbb{E}_{\pi_\theta}[f(s,a)] \approx \frac{1}{N}\sum_{n=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\nabla_\theta \log \pi_\theta(a_t^n|s_t^n)Q^{\pi_\theta}(s_t^n, a_t^n)$$

各要素の意味

<div style="page-break-after: always;"></div>

交差エントロピー誤差

$$E = -\frac{1}{N}\sum_{n}\sum_{k} t_{nk} \log y_{nk}$$

各要素の意味

- N : データ個数
- k : データの次元数

<div style="page-break-after: always;"></div>

アフィンレイヤ

$$H = XW + B$$

$$\frac{\partial L}{\partial W} = X^T \frac{\partial L}{\partial H}$$

各要素の意味

<div style="page-break-after: always;"></div>

モーメンタム

$$v_{t+1} = av_t - \eta \frac{\partial L}{\partial \theta_t}$$

$$\theta_{t+1} = \theta_t + v_{t+1}$$

各要素の意味

<div style="page-break-after: always;"></div>

確率的勾配降下法　※1/8追加

各要素の意味

<div style="page-break-after: always;"></div>

NesterovAG

$$v_{t+1} = av_t - \eta \frac{\partial L}{\partial(\theta_t + av_t)}$$

$$\theta_{t+1} = \theta_t + v_{t+1}$$

各要素の意味

<div style="page-break-after: always;"></div>

AdaGrad

$$h_{t+1} = h_t + \frac{\partial L}{\partial \theta_t} \odot \frac{\partial L}{\partial \theta_t}$$

$$\theta_{t+1} = \theta_t - \eta \frac{1}{\varepsilon + \sqrt{h_{t+1}}} \odot \frac{\partial L}{\partial \theta_t}$$

各要素の意味

<div style="page-break-after: always;"></div>

RMSProp

$$h_{t+1} = \rho h_t + (1-\rho)\frac{\partial L}{\partial \theta_t} \odot \frac{\partial L}{\partial \theta_t}$$

$$\theta_{t+1} = \theta_t - \eta \frac{1}{\sqrt{\varepsilon + h_{t+1}}} \odot \frac{\partial L}{\partial \theta_t}$$

各要素の意味

<div style="page-break-after: always;"></div>

Adam

$$m_{t+1} = \rho_1 m_t + (1 - \rho_1) \frac{\partial L}{\partial \theta_t}$$

$$v_{t+1} = \rho_2 v_t + (1 - \rho_2) \frac{\partial L}{\partial \theta_t} \odot \frac{\partial L}{\partial \theta_t}$$

$$\hat{m}_{t+1} = \frac{m_{t+1}}{1 - \rho_1^t}$$

$$\hat{v}_{t+1} = \frac{v_{t+1}}{1 - \rho_2^t}$$

$$\theta_{t+1} = \theta_t - \eta \frac{1}{\sqrt{\quad}} \odot \hat{m}_{t+1}$$

バッチ正規化

$$h' = \frac{h - \mu}{\sigma}$$
$$\gamma h' + \beta$$

各要素の意味

<div style="page-break-after: always;"></div>

畳み込み

$$(I * K)(i,j) = \sum_m \sum_n I(i+m, j+n)K(m,n)$$

各要素の意味

<div style="page-break-after: always;"></div>

IoU

$$IoU(B_{true}, B_{pred}) = \frac{|B_{true} \cap B_{pred}|}{|B_{true} \cup B_{pred}|} = \frac{|B_{true} \cap B_{pred}|}{|B_{true}| + |B_{pred}| - |B_{true} \cap B_{pred}|}$$

各要素の意味

<div style="page-break-after: always;"></div>

Dice 係数

$$Dice(S_{true}, S_{pred}) = \frac{|S_{true} \cap S_{pred}|}{\frac{|S_{true}|+|S_{pred}|}{2}} = \frac{2|S_{true} \cap S_{pred}|}{|S_{true}| + |S_{pred}|}$$

各要素の意味

<div style="page-break-after: always;"></div>

AP

$$AP = \frac{1}{11} \sum_{r \in \{0,0.1,0.2,\ldots,1\}} p_{interp}(r)$$

$$p_{interp}(r) = \max_{\tilde{r} \geq r} p(\tilde{r})$$

各要素の意味

<div style="page-break-after: always;"></div>

LSTMの順伝播

$$G = \tanh\left(X_t W_x^{(g)} + H_{t-1} W_h^{(g)} + b^{(g)}\right)$$

$$I = \mathrm{sigmoid}\left(X_t W_x^{(i)} + H_{t-1} W_h^{(i)} + b^{(i)}\right)$$

$$F = \mathrm{sigmoid}\left(X_t W_x^{(f)} + H_{t-1} W_h^{(f)} + b^{(f)}\right)$$

$$O = \mathrm{sigmoid}\left(X_t W_x^{(o)} + H_{t-1} W_h^{(o)} + b^{(o)}\right)$$

各要素の意味

<div style="page-break-after: always;"></div>

GRUの順伝播

$$R = \mathrm{sigmoid}\left(X_t W_x^{(r)} + H_{t-1} W_h^{(r)} + b^{(r)}\right)$$

$$Z = \mathrm{sigmoid}\left(X_t W_x^{(z)} + H_{t-1} W_h^{(z)} + b^{(z)}\right)$$

$$\tilde{H} = \tanh\left\{X_t W_x^{(\tilde{h})} + (R \odot H_{t-1}) W_h^{(\tilde{h})} + b^{(\tilde{h})}\right\}$$

$$H_t = Z \odot H_{t-1} + (1 - Z) \odot \tilde{H}$$

各要素の意味

<div style="page-break-after: always;"></div>

WaveNetの定式化

$$p(x) = \prod_{t=1}^{T} p(x_t | x_1, x_2, \ldots, x_{t-1})$$

各要素の意味

<div style="page-break-after: always;"></div>

TransformerのScaled Dot-Product Attention

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

各要素の意味

<div style="page-break-after: always;"></div>

TransformerのPositional Encoding

$$PE_(pos, 2i) = \sin\left(pos/10000^{2i/d_{model}}\right)$$

$$PE_(pos, 2i+1) = \cos\left(pos/10000^{2i/d_{model}}\right)$$

各要素の意味

<div style="page-break-after: always;"></div>

VAEの損失関数

$$-\log p(x) \leq -L = \mathbb{E}_{z \sim p(z|x)}[-\log p(x|z)] + \int \log \left(\frac{p(x|z)}{p(z)} \right) p(z|x) dz$$

各要素の意味

<div style="page-break-after: always;"></div>

GANの定式化

$$\min_{G} \max_{D} \mathbb{E}_{\mathbb{x}} [\log D(x)] + \mathbb{E}_{\mathbb{z}} [\log(1 - D(G(z)))]$$

各要素の意味

<div style="page-break-after: always;"></div>

DQN

$$L(\theta) = \mathbb{E}_{\mathbb{s},\mathbb{a},\mathbb{r},\mathbb{s}'\sim\mathbb{D}}\left[\left(r + \gamma \max_{a'} Q(s', a'; \theta^-) - Q(s, a; \theta)\right)^2\right]$$

各要素の意味

<div style="page-break-after: always;"></div>

蒸留における温度付きソフトマックス関数

$$\mathrm{Softmax}(z)_i = \frac{\exp(z_i/T)}{\Sigma_j \exp(z_j/T)}$$

各要素の意味