### 数式を中心に復習する

### 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

### 特異値分解

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

#### 各要素の意味

- A は
- *U* は
- Σ は
- ullet  $V^T$  (t

### 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) ( $\ddagger$
- $p^x(1-p)^{1-x}$  (‡

### ベルヌーイ分布の期待値

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

$$= p$$

### 各要素の意味

### ベルヌーイ分布の分散

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

#### 各要素の意味

### マルチヌーイ分布(カテゴリ 分布)

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

#### 各要素の意味

### マルチヌーイ分布(カテゴリ 分布)の負の対数尤度

$$-\log L_D(p) = -\log \prod_{i=1}^n f(x_i;p)$$

$$=-\sum_{i=1}^n\log\prod_{j=1}^k p_j^{x_{ij}}$$

$$=-\sum_{i=1}^n\sum_{j=1}^k\log p_j^{x_{ij}}$$

$$x_i = -\sum_{i=1}^n \sum_{j=1}^n x_{ij} \log p_j$$

### 正規分布

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

#### 各要素の意味

### 正規分布の負の対数尤度

$$L(\mu) = \prod_{i=1}^{n} f(x_i; \mu)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_i - \mu)^2\right)$$

$$-\log L(\mu) = -\log\left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_i - \mu)^2\right)\right)$$

$$= -\sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_i - \mu)^2\right)\right)$$

$$= -\sum_{i=1}^{n} \left(\log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{1}{2}(x_i - \mu)^2\right)$$

$$= -n\log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{1}{2}\sum_{i=1}^{n} (x_i - \mu)^2$$

### 正規分布の最尤推定

$$egin{align} rac{d}{d\mu}g(\mu) &= rac{1}{2} \sum_{i=1}^n rac{d}{d\mu} (x_i - \mu)^2 \ &= rac{1}{2} \sum_{i=1}^n (-2(x_i - \mu)) \ &= \sum_{i=1}^n \mu - \sum_{i=1}^n x_i \ &= n\mu - \sum_{i=1}^n x_i \ \hat{\mu} &= rac{1}{2} \sum_{i=1}^n x_i \ \end{pmatrix}$$

### エントロピー

$$H(X) = -\sum_x p(x) \log_2 p(x)$$

#### 各要素の意味

# 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

## 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

### KLダイバージェンス

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

#### 各要素の意味

### JSダイバージェンス

$$D_{JS}(p||q) = rac{1}{2}igg(\sum_x p(x)\log_2rac{p(x)}{r(x)} + \sum_x q(x)\log_2rac{q(x)}{r(x)}igg) \ r(x) = rac{p(x)+q(x)}{2}$$

#### 各要素の意味

### ベイズの定理

$$p(C|x) = rac{p(x|C)p(C)}{p(x)}$$

#### 各要素の意味

### バイアス・バリアンス・ノイ ズ

$$egin{aligned} \mathbb{E}(L) &= \int \{y(x) - h(x)\}^2 p(x) dx + \iint \{h(x) - t\}^2 p(x,t) dx \ &\int \{\mathbb{E}_D[y(x;D)] - h(x)\}^2 p(x) dx \ &\int \mathbb{E}_D[\{y(x;D) - \mathbb{E}_D[y(x;D)]\}^2 p(x) dx \ &\iint \{h(x) - t\}^2 p(x,t) dx dt \end{aligned}$$

#### 各要素の意味

### 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、 $f(x)=rac{1}{2}$ となる

#### ReLU

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

#### 各要素の意味

### オッズ

$$egin{aligned} rac{p(y=1|x)}{p(y=0|x)} &= rac{\hat{y}}{1-\hat{y}} \ rac{\hat{y}}{1-\hat{y}} &= rac{rac{1}{1+\exp(-w^Tx-b)}}{1-rac{1}{1+\exp(-w^Tx-b)}} \ &= rac{1}{(1+\exp(-w^Tx-b))-1} \ &= rac{1}{\exp(-w^Tx-b)} \ &= \exp(w^Tx-b) \end{aligned}$$

### ガウスカーネル

$$k(x,x') = \exp\left(-rac{||x-x'||^2}{eta}
ight)$$

#### 各要素の意味

### 正則化

$$egin{aligned} E + \lambda_2 ||w||_2^2 \ E + \lambda_1 ||w||_1 \ E + \lambda_1 ||w||_1 + \lambda_2 ||w||_2^2 \end{aligned}$$

#### 各要素の意味

### ソフトマックス

$$ext{softmax}(z)_i = rac{\exp(z_i)}{\Sigma_j \exp(z_j)}$$

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

#### 各要素の意味

### 二乗和誤差

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

### 各要素の意味

### 生成モデル

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

#### 各要素の意味

### ベルマン方程式

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

### 各要素の意味

#### **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

### Q学習

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

#### 各要素の意味

### 方策勾配定理

$$egin{aligned} 
abla_{ heta} J( heta) &= \sum_{d} d^{\pi_{ heta}}(s) \sum_{d} 
abla_{ heta} \pi_{ heta}(a|s, heta) Q^{\pi_{ heta}}(s,a) \ 
abla_{ heta} \log \pi_{ heta}(a|s) &= rac{\partial \pi_{ heta}(a|s)}{\partial heta} rac{1}{\pi_{ heta}(a|s)} \ 
d^{\pi_{ heta}}(s) &= \sum_{k=0}^{\infty} \gamma^k P^{\pi_{ heta}}(s_k = s|s_0) \ 
abla_{ heta} J( heta) &= \sum_{d} d^{\pi_{ heta}}(s) \sum_{a} (
abla_{ heta}(a|s, heta)) Q^{\pi_{ heta}}(s,a) \ 
&= \sum_{d} d^{\pi_{ heta}}(s) \sum_{a} \pi_{ heta}(a|s, heta) (
abla_{ heta} \log \pi_{ heta}(a|s, heta)) Q^{\pi_{ heta}}(s,a) \ 
&= \mathbb{E}_{\pi_{ heta}}[
abla_{ heta} \log \pi_{ heta}(a|s, heta) Q^{\pi_{ heta}}(s,a)] \end{aligned}$$

#### 各要素の意味

### モンテカルロ近似

$$egin{aligned} 
abla_{ heta}J( heta) &= \mathbb{E}_{\pi_{ heta}}[f(s,a)] pprox rac{1}{N} \sum_{n=1}^{N} rac{1}{T} \sum_{t=1}^{T} 
abla_{ heta} \log \pi_{ heta}(a^n_t|s^n_t) Q^{\pi_{ heta}}(s^n_t,a^n_t) \end{aligned}$$

#### 各要素の意味

### 交差エントロピー誤差

$$E = -rac{1}{N} \sum_n \sum_k t_{nk} \log y_{nk}$$

#### 各要素の意味

- N:データ個数
- *k*: データの次元数

### アフィンレイヤ

$$\frac{H = XW + B}{\partial L} = X^T \frac{\partial L}{\partial H}$$

#### 各要素の意味

### モーメンタム

$$egin{aligned} v_{t+1} &= a v_t - \eta rac{\partial L}{\partial heta_t} \ heta_{t+1} &= heta_t + v_{t+1} \end{aligned}$$

#### 各要素の意味

### 確率的勾配降下法 ※1/8追加

#### 各要素の意味

#### **NesterovAG**

$$egin{aligned} v_{t+1} &= av_t - \eta rac{\partial L}{\partial ( heta_t + av_t)} \ heta_{t+1} &= heta_t + v_{t+1} \end{aligned}$$

#### 各要素の意味

#### AdaGrad

$$egin{align} h_{t+1} &= h_t + rac{\partial L}{\partial heta_t} \odot rac{\partial L}{\partial heta_t} \ heta_{t+1} &= heta_t - \eta rac{1}{arepsilon + \sqrt{h_{t+1}}} \odot rac{\partial L}{\partial heta_t} \ \end{matrix}$$

#### 各要素の意味

## **RMSProp**

$$egin{align} h_{t+1} &= 
ho h_t + (1-
ho) rac{\partial L}{\partial heta_t} \odot rac{\partial L}{\partial heta_t} \ &= heta_t - \eta rac{1}{\sqrt{arepsilon + h_{t+1}}} \odot rac{\partial L}{\partial heta_t} \ &= rac{\partial L$$

#### 各要素の意味

#### Adam

$$egin{align} m_{t+1} &= 
ho_1 m_t + (1-
ho_1) rac{\partial L}{\partial heta_t} \ v_{t+1} &= 
ho_2 v_t + (1-
ho_2) rac{\partial L}{\partial heta_t} \odot rac{\partial L}{\partial heta_t} \end{array}$$

$$\hat{m}_{t+1} = rac{m_{t+1}}{1-
ho_1^t} \ \hat{v}_{t+1} = rac{v_{t+1}}{1-
ho_2^t}$$

$$heta_{t+1} = heta_t - \eta$$
  $\hat{m}_{t+1}$ 

# バッチ正規化

$$h' = rac{h - \mu}{\sigma}$$
  $\gamma h' + eta$ 

#### 各要素の意味

## 畳み込み

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

#### 各要素の意味

#### IoU

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

#### 各要素の意味

## Dice 係数

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

#### 各要素の意味

#### AP

$$AP = rac{1}{11} \sum_{r \in \{0,0.1,0.2,...,1\}} p_{interp}(r) \ p_{interp}(r) = \max_{ ilde{r} \geq r} p( ilde{r})$$

#### 各要素の意味

## LSTMの順伝播

$$egin{aligned} G &= anh \left( X_t W_x^{(g)} + H_{t-1} W_h^{(g)} + b^{(g)} 
ight) \ I &= ext{sigmoid} \left( X_t W_x^{(i)} + H_{t-1} W_h^{(i)} + b^{(i)} 
ight) \ F &= ext{sigmoid} \left( X_t W_x^{(f)} + H_{t-1} W_h^{(f)} + b^{(f)} 
ight) \ O &= ext{sigmoid} \left( X_t W_x^{(o)} + H_{t-1} W_h^{(o)} + b^{(o)} 
ight) \end{aligned}$$

#### 各要素の意味

## GRUの順伝播

$$egin{aligned} R &= \operatorname{sigmoid}\left(X_tW_x^{(r)} + H_{t-1}W_h^{(r)} + b^{(r)}
ight) \ Z &= \operatorname{sigmoid}\left(X_tW_x^{(z)} + H_{t-1}W_h^{(z)} + b^{(z)}
ight) \end{aligned}$$

$$ilde{H} = anh \left\{ X_t W_x^{( ilde{h})} + (R \odot H_{t-1}) W_h^{( ilde{h})} + b^{( ilde{h})} 
ight\}$$

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

#### 各要素の意味

## WaveNetの定式化

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

#### 各要素の意味

# Transformer Scaled Dot-Product Attention

$$\operatorname{Attention}(Q,K,V) = \operatorname{softmax}\left(rac{QK^T}{\sqrt{d_k}}
ight)V$$

#### 各要素の意味

# Transformer Positional Encoding

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

#### 各要素の意味

## VAEの損失関数

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

#### 各要素の意味

## GANの定式化

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

#### 各要素の意味

#### **DQN**

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

#### 各要素の意味

# 蒸留における温度付きソフト マックス関数

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

#### 各要素の意味