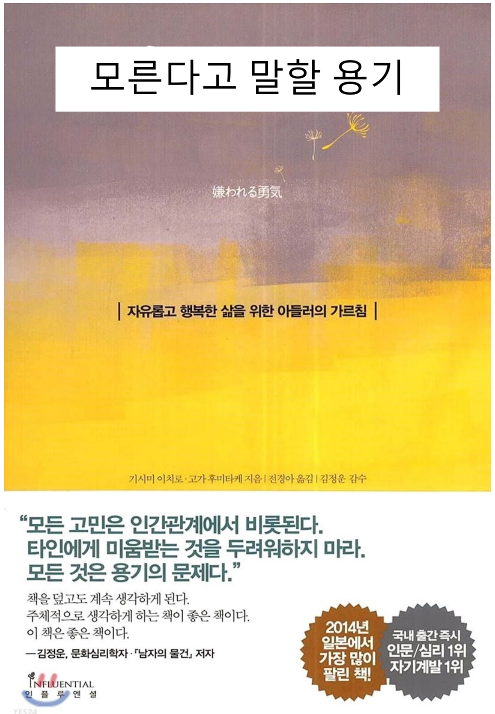
Dealing uncertainty
모른다고 말할 용기

김기훈

Vision and Display System Lab.
Sogang University

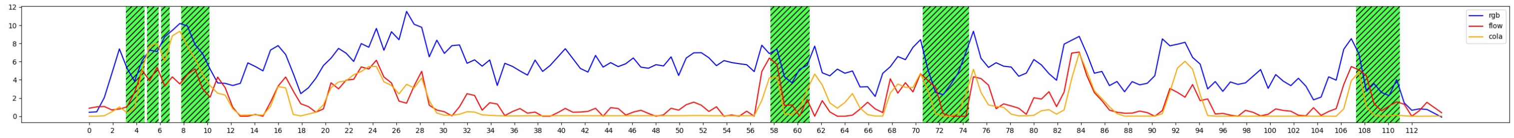
Outline

- Motivation
- What is EDL?
- What is Open Set?
- Paper
- Appendix
- Reference



Motivation

- Weakly Supervised Learning
  - Original task: WSTAL (Weakly Supervised Temporal Action Localization)
    - Label: Classes in the given video
    - Task: To classify & localize actions
  - Small proportion taken by actions



  - Start, end time stamps are subjective and inconsistent

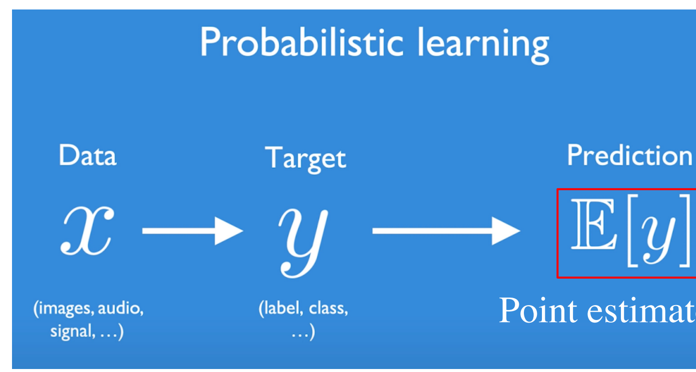
What is EDL?

- Motivation

Model

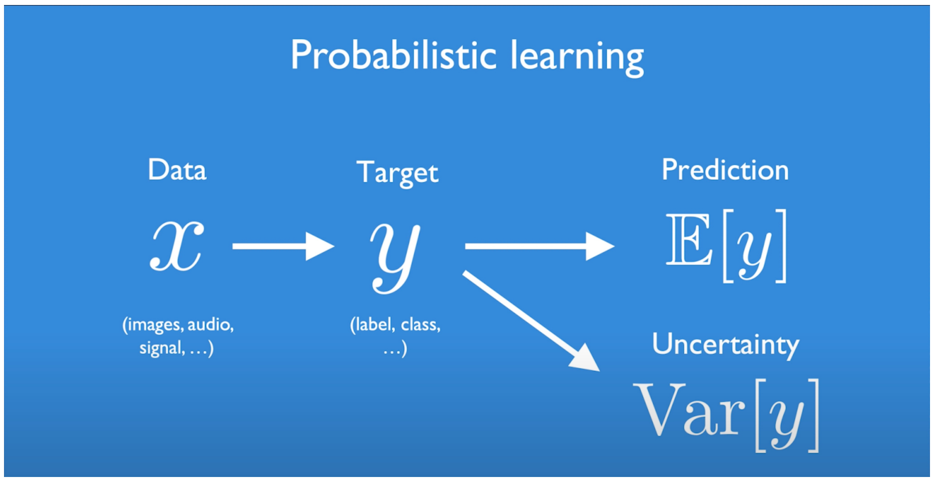
Reality

- Probabilisic Model
  - 현재 사용되는 가장 기본적인 형태의 모델

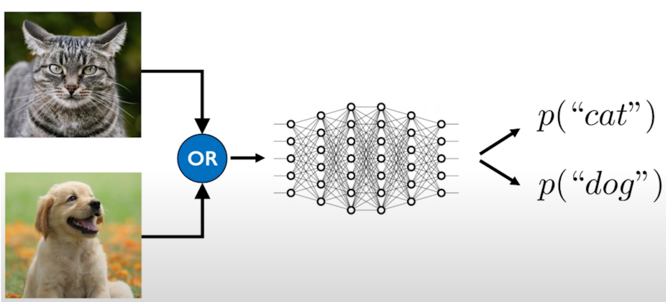


What is EDL?

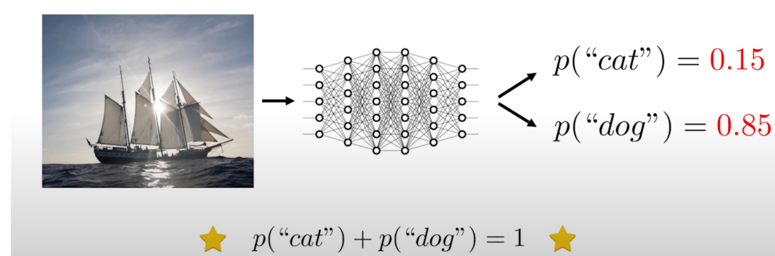
- Uncertainty



- Likelihood vs. Confidence



The output likelihoods will be unreliable if the input is ***unlike anything during training***



What is EDL?

- Mainly from "Deep Evidential Regression" (NeurIPS 2020)
  - 세상에는 2가지의 uncertainty가 있다고 가정
    - Epistemic
    - Aleatoric
- Modeling uncertainties
  - Regression에서의 흔한 setting
    - Uncertainty에 대하여 어떤 정보도 주지 않음!
  - MLE approach
    - Negative log likelihood loss function을 minimizing하는 것과 같은 결과
    - Epistemic uncertainty에 대해서 어떤 정보도 주지 않음!

$$\min_{\boldsymbol{w}} J(\boldsymbol{w}); \quad J(\boldsymbol{w}) = \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}_i(\boldsymbol{w})$$

$$\mathcal{L}_i(\boldsymbol{w}) = -\log p(y_i | \underbrace{\mu, \sigma^2}_{\boldsymbol{\theta}}) = \frac{1}{2}\log(2\pi\sigma^2) + \frac{(y_i - \mu)^2}{2\sigma^2}.$$

서강대학교 SOGANG UNIVERSITY

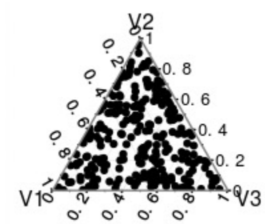
VDS LAB

Preliminaries

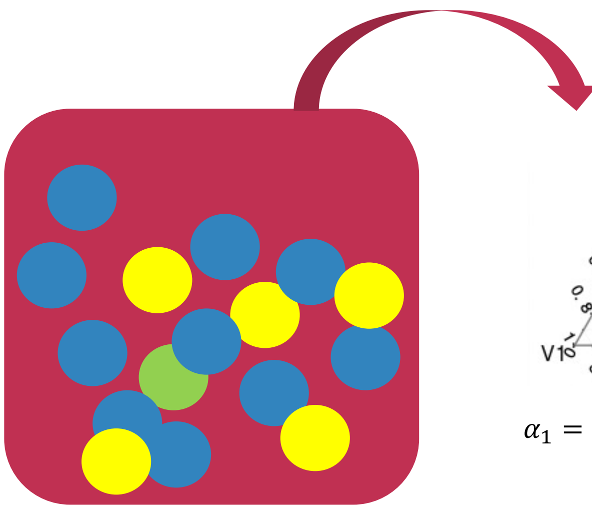
- Statistics
  - Dirichlet distribution
  - Conjugate priors

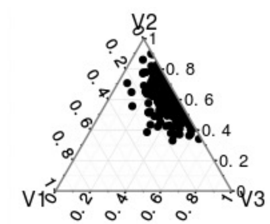
Dirichlet distribution

- Definition: Multivariate probability distribution, parameterized by a vector of positive-real value parameters $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_k)$
  - Higher value of α_i, the greater "weight" of X_i
- What does this distribution models?



$\alpha_1 = \alpha_2 = \alpha_3 = 1$

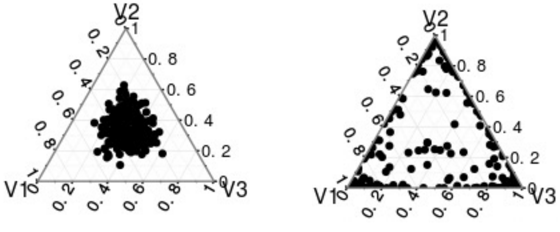




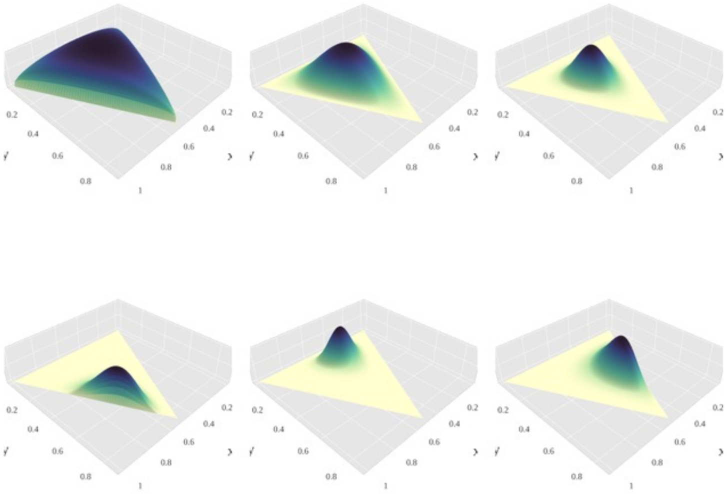
$\alpha_1 = 1, \alpha_2 = 10, \alpha_3 = 5$

Dirichlet Distribution

- Modeling
  - (a) $\alpha_1 = \alpha_2 = \alpha_3 = 10$
  - (b) $\alpha_1 = \alpha_2 = \alpha_3 = 0.2$
    - $\alpha < 1$: Pushes x_i towards extremes
    - $\alpha > 1$: Attracts x_i toward central value

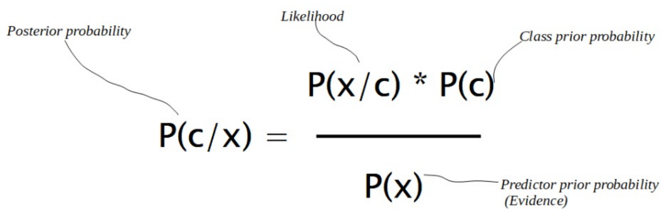


(a) (b)



Conjugate Prior

- Definition: Likelihood와 Prior로 구한 Posterior의 분포가 Prior의 분포와 같아지게하는 Prior를 지칭



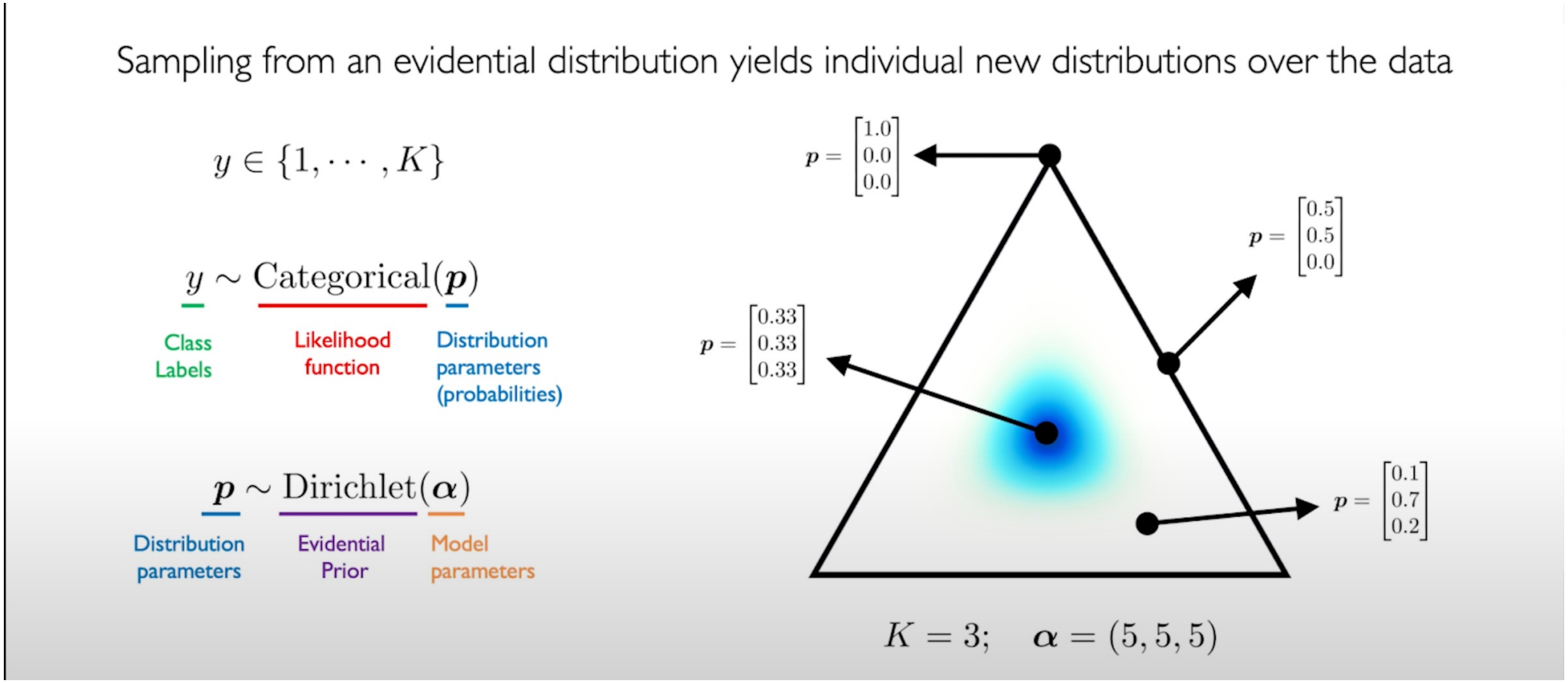
$P(c/x) = P(x_1/c) * P(x_2/c) * \ldots P(x_n/c) * P(c)/P(x)$

Likelihood	Model parameters	Conjugate prior distribution	Prior hyperparameters	Posterior hyperparameters[note 1]	Interpretation of hyperparameters	Posterior predictive[note 2]
Bernoulli	p (probability)	Beta	$\alpha,\ \beta \in \mathbb{R}$	$\alpha + \sum_{i=1}^n x_i,\ \beta + n - \sum_{i=1}^n x_i$	α successes, β failures[note 3]	$p(\tilde{x}=1) = \frac{\alpha'}{\alpha'+\beta'}$
Binomial	p (probability)	Beta	$\alpha,\ \beta \in \mathbb{R}$	$\alpha + \sum_{i=1}^n x_i,\ \beta + \sum_{i=1}^n N_i - \sum_{i=1}^n x_i$	α successes, β failures[note 3]	$\operatorname{BetaBin}(\tilde{x}\mid\alpha',\beta')$ (beta-binomial)
Negative binomial with known failure number, r	p (probability)	Beta	$\alpha,\ \beta \in \mathbb{R}$	$\alpha + rn,\ \beta + \sum_{i=1}^n x_i$	α total successes, β failures[note 3] (i.e., $\frac{\beta}{r}$ experiments, assuming r stays fixed)	$\operatorname{BetaNegBin}(\tilde{x}\mid\alpha',\beta')$ (beta-negative binomial)
Poisson	λ (rate)	Gamma	$k,\ \theta \in \mathbb{R}$	$k + \sum_{i=1}^n x_i,\ \frac{\theta}{n\theta+1}$	k total occurrences in $\frac{1}{\theta}$ intervals	$\operatorname{NB}\left(\tilde{x} \mid k', \frac{\theta'}{\theta'+1}\right)$ (negative binomial)
			$\alpha,\ \beta$[note 4]	$\alpha + \sum_{i=1}^n x_i,\ \beta + n$	α total occurrences in β intervals	$\operatorname{NB}\left(\tilde{x} \mid \alpha', \frac{1}{1+\beta'}\right)$ (negative binomial)
Categorical	$\boldsymbol{p}$ (probability vector), k (number of categories; i.e., size of $\boldsymbol{p}$)	Dirichlet	$\boldsymbol{\alpha} \in \mathbb{R}^k$	$\boldsymbol{\alpha} + (c_1, \ldots, c_k)$, where c_i is the number of observations in category i	α_i occurrences of category i[note 3]	$p(\tilde{x}=i) = \frac{{\alpha_i}'}{\sum_i {\alpha_i}'} = \frac{\alpha_i + c_i}{\sum_i \alpha_i + n}$
Multinomial	$\boldsymbol{p}$ (probability vector), k (number of categories; i.e., size of $\boldsymbol{p}$)	Dirichlet	$\boldsymbol{\alpha} \in \mathbb{R}^k$	$\boldsymbol{\alpha} + \sum_{i=1}^n \mathbf{x}_i$	α_i occurrences of category i[note 3]	$\operatorname{DirMult}(\tilde{\mathbf{x}} \mid \boldsymbol{\alpha}')$ (Dirichlet-multinomial)
Hypergeometric with known total population size, N	M (number of target members)	Beta-binomial[3]	$n = N, \alpha,\ \beta$	$\alpha + \sum_{i=1}^n x_i,\ \beta + \sum_{i=1}^n N_i - \sum_{i=1}^n x_i$	α successes, β failures[note 3]	
Geometric	p_0 (probability)	Beta	$\alpha,\ \beta \in \mathbb{R}$	$\alpha + n,\ \beta + \sum_{i=1}^n x_i$	α experiments, β total failures[note 3]	

- 대표적인 예시
  - Binomial likelihood + Beta prior = Beta posterior
  - Poisson likelihood + Gamma prior = Gamma posterior
  - Normal likelihood + Inv-Gamma prior = Inv-Gamma posterior
- Why use conjugate prior?
  - Prior와 likelihood가 같은 algebraic form을 갖는다면, posterior를 쉽게 구할 수 있게 된다

Conjugate Prior

- Conjugate Prior를 이용해서 p를 model이 찾게하면, Categorical likelihood와 짝을 이뤄, y도 Dirichlet distribution의 형태를 갖게 할 수 있다



Learning the Evidential Distribution

- Subjective opinion
  - $\omega = (\boldsymbol{b}, u, \boldsymbol{a})$
    - b: belief mass
  - $p_k = b_k + a_k u$
    - $a_k = \frac{1}{K}$
    - a_k는 base rate로 uncertainty에 얼만큼의 weight를 주는지에 대한 정보
  - p_k는 Dirichlet distribution (9)를 따른다고 가정
  - Evidence와 Dirichlet strength α는 식 (10)과 같은 관계를 갖는다

$$u + \sum_{k=1}^{K} b_k = 1 \tag{8}$$

$$\mathrm{Dir}(\mathbf{p}|\boldsymbol{\alpha}) = \begin{cases} \frac{1}{B(\boldsymbol{\alpha})} \prod_{k=1}^{K} p_k^{\alpha_k - 1}, & \text{for } \mathbf{p} \in \mathcal{S}_K, \\ 0, & \text{otherwise,} \end{cases} \tag{9}$$

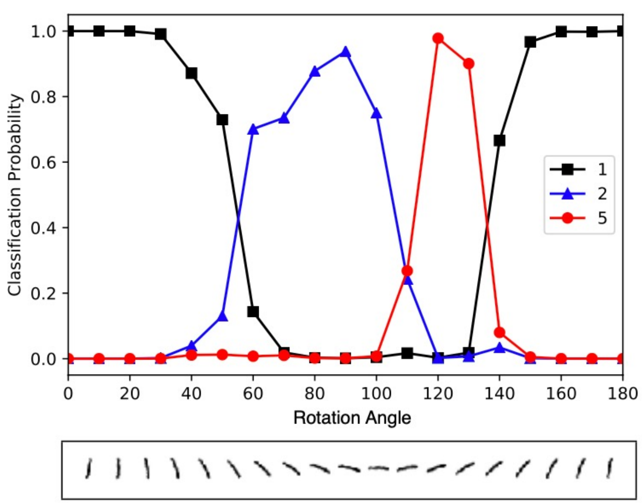
$$\boldsymbol{\alpha} = \mathbf{e} + \mathbf{a}W \tag{10}$$

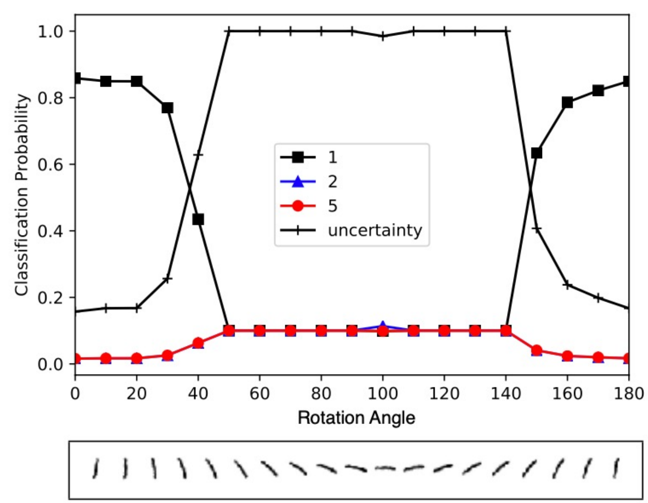
Learning the Evidential Distribution

- W를 K로 설정해도 아무런 문제가 없으며, $a_k = \frac{1}{K}$이기 때문에 식 (10)으로부터 다음과 같은 관계가 성립
  - $a_k = e_k + 1$
- 정리하면 Dirichlet evidence는 아래와 같이 정리할 수 있게 된다
  - $b_k = \frac{e_k}{S}$
  - $u = \frac{K}{S}$

$$\boldsymbol{\alpha} = \mathbf{e} + \mathbf{a}W \tag{10}$$

$$\mathcal{L}^{(i)}_{EDL}(\mathbf{y}, \mathbf{e}; \theta) = -\log\left(\int \prod_{k=1}^{K} p_{ik}^{y_{ik}} \frac{1}{B(\boldsymbol{\alpha}_i)} \prod_{k=1}^{K} p_{ik}^{\alpha_{ik}-1} d\mathbf{p}_i\right)$$
$$= \sum_{k=1}^{K} y_{ik}\left(\log(S_i) - \log(e_{ik}+1)\right) \tag{13}$$





What is Open Set Recognition?

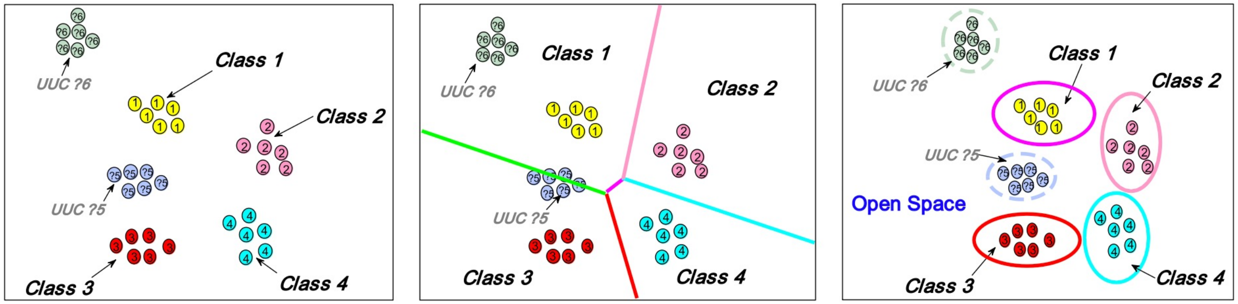
- Modeling the world with 4 categories

Side-information

To Model

	Known	Unknown
Known	KKC (Known Known Classes)	KUC (Known Unknown Classes)
Unknown	UKC (Unknown Known Classes)	UUC (Unknown Unknown Classes)

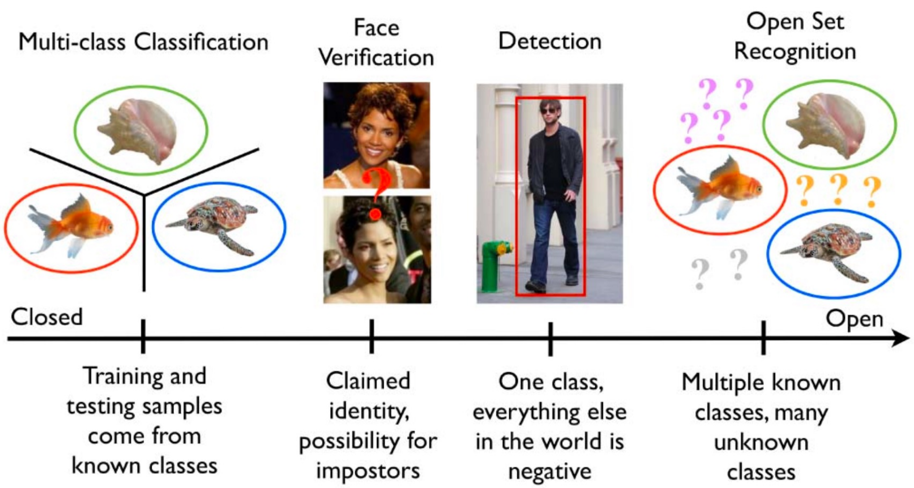
- Open set recognition scenario
  - UUCs appear in testing
  - How will model deal with UUC?
    - Models should be able to ”reject”



(a) Distribution of the original data set. (b) Traditional recognition/classification problem. (c) Open set recognition/classification problem.

What is Open Set Recognition

- Openness
  - Definition
    - $O = 1 - \sqrt{\frac{2 \times |C_{TR}|}{|C_{TA}| + |C_{TE}|}}$ (in [1])
    - TR: Training classes, TA: Target classes, TE: Testing classes
    - Relationship in most works
    - TA: classes to be recognized
  - Modified definition
    - $O^* = 1 - \sqrt{\frac{2 \times |C_{TR}|}{|C_{TR}| + |C_{TE}|}}$ (in [2])



Forumulation of OSR

- Formulation
  - Open space risk
    - 전체 space에서의 정답과 open space에서의 정답의 비

$$R_{\mathcal{O}}(f) = \frac{\int_{\mathcal{O}} f(x)dx}{\int_{S_o} f(x)dx}$$

open space

Open space + positive training examples

  - Open set recognition problem
    - Open space risk와 empirical risk를 최소화하는 function f를 찾는 것

$$\arg\min_{f \in \mathcal{H}} \{R_{\mathcal{O}}(f) + \lambda_r R_{\varepsilon}(f(V))\}$$

Main Paper

- "Evidential Deep Learning for Open Set Action Recognition"
  - ICCV 2021 (oral)
  - Dealing action recognition + open set + evidence deep learning

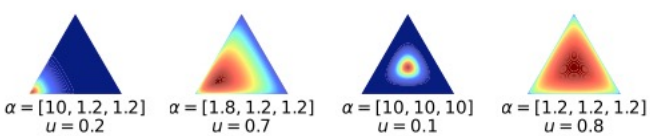
- Motivation
  - Real-world scenario most of human action are out of distribution from training data

- Contribution
  - Uncertainty evaluation을 통한 open set action recognition
  - Overconfident prediction, static bias 문제를 해결
    - EUC (Evidential Uncertainty Calibration)
    - CED (Contrastive Evidence Debiasing)
  - 본 논문에서 제안한 DEAR 방법이 SOTA action recognition의 성능을 boost

EUC

- Evidential uncertainty calibration
  - Problem: NLL(negative log-likelihood)은 over-fit가 발생할 확률이 높다
  - AvU를 Maximize하는것을 목표로 학습시킨다
    - AU와 IC는 최대한 적게 나오도록 학습
    - Uncertainty와 accuracy 사이의 consistency를 학습

$$\text{AvU} = \frac{n_{AC} + n_{IU}}{n_{AC} + n_{AU} + n_{IC} + n_{IU}}$$



(a) AC (b) AU (c) IC (d) IU

(a) 정확하고 확실한 경우
(b) 정확하지만 확실하지 않은 경우
(c) 정확하지 않지만, 확실한 경우
(d) 정확하지도 않고 확실하지도 않은 경우

$$\mathcal{L}_{EUC} = -\lambda_t \sum_{i \in \{\hat{y}_i = y_i\}} p_i \log(1 - u_i) - (1 - \lambda_t) \sum_{i \in \{\hat{y}_i \neq y_i\}} (1 - p_i) \log(u_i)$$

When match, accuracy → 1, uncertainty → 0
Else, accuracy → 0, uncertainty → 1
λ_t: $annealing\ fator$

CED

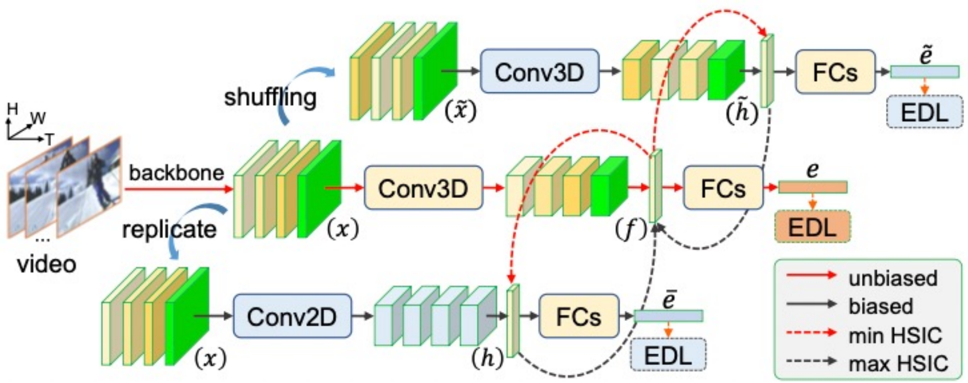
- Contastive evidence debiasing
  - Problem: Static bias
    - Scene bias
      - Basketball dunk → Basketball court
    - Object bias
      - Playing piano → Piano
    - Human bias
      - Brushing hair, Military marching

  - Model recognition based to the background of the water and sky → When those are gone?!
    - Mimetics dataset (by Naver Labs Europe)
      - 713 video clips from Youtube
      - Subset of 50 classes from Kinetics400 dataset

CED (Cont.)

- How to solve "evidence debiasing"

$$\mathrm{HSIC}^{k,l}(U,V) = \frac{1}{m(m-3)}\left[\mathrm{tr}(\tilde{U}\tilde{V}^T) + \frac{\mathbf{1}^T\tilde{U}\mathbf{1}\mathbf{1}^T\tilde{V}\mathbf{1}}{(m-1)(m-2)} - \frac{2}{m-2}\mathbf{1}^T\tilde{U}\tilde{V}^T\mathbf{1}\right]$$



$$HSIC^{k_1,k_2}(f,h) = 0$$

iff h and f is independent

- HSIC

$$\mathcal{L}(\theta_f,\phi_f) = \mathcal{L}_{EDL}(\mathbf{y},\mathbf{e};\theta_f,\phi_f) + \lambda\sum_{\mathbf{h}\in\Omega}\mathrm{HSIC}(\mathbf{f},\mathbf{h};\theta_f)$$

f to be independent of the biased feature **h**

$$\Omega = \{h_{3D}(\tilde{\mathbf{x}}), h_{2D}(\mathbf{x})\}$$

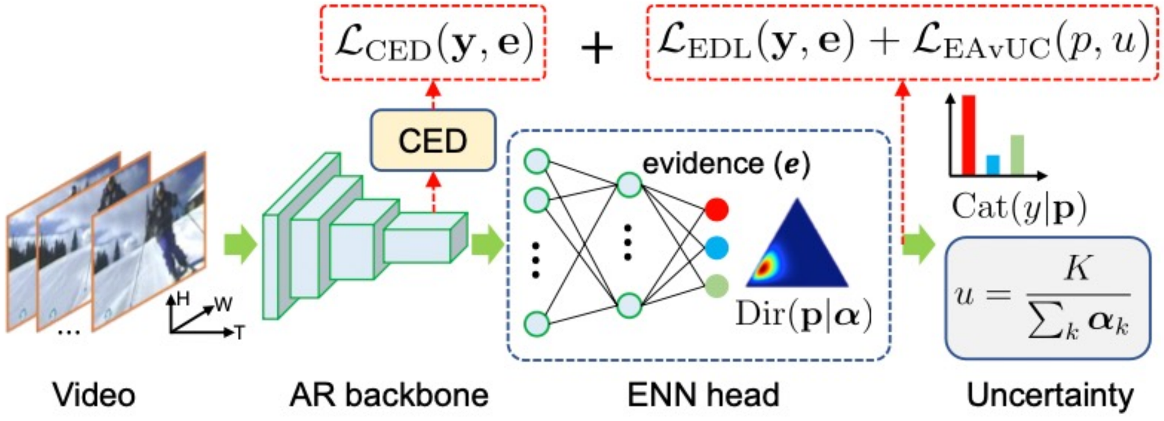
$$\mathcal{L}(\theta_h,\phi_h) = \sum_{\mathbf{h}\in\Omega}\{\mathcal{L}_{EDL}(\mathbf{y},\mathbf{e}_h;\theta_h,\phi_h) - \lambda\mathrm{HSIC}(\mathbf{f},\mathbf{h};\theta_h)\}$$

To avoid the biased feature h to predict arbitrary evidence

h is similar enough to f

Model Architecture (DEAR)

- Proposed DEAR Method
  - 기존의 AR backbone을 사용
  - ENN head에서는 evidence를 구함
  - 최종적으로 구한 uncertainty(u)를 가지고 unknown을 reject



Experiments

- Dataset
  - UCF-101 , HMDB-51 , MiT-v2

	#Videos	#Categories	Year	Comments
UCF-101	13,320	101	2012	5 sub-categories
HMDB-51	6,766	51	2011	
MiT-v2	30,500	305	2019	Categories are from WordNet

  - Models are trained on UCF-101
    - UCF-101의 class가 closed set을 의미
    - Openness 상황을 위해서 i개의 new class가 나머지 dataset에서 임의 선택

Experiments

- Evaluation Protocol
  - Closed set accuracy (for K-class classification)
  - Open set accuracy
    - 2 classes (known, unknown)
      - Open Set area under ROC curve
    - (K+1) classes (known k classes and unknown class)
      - Open maF1(macro-F1 score)

$$\text{Open maF1} = \frac{\sum_i \omega_O^{(i)} \cdot F_1^{(i)}}{\sum_i \omega_O^{(i)}}$$

Experiments

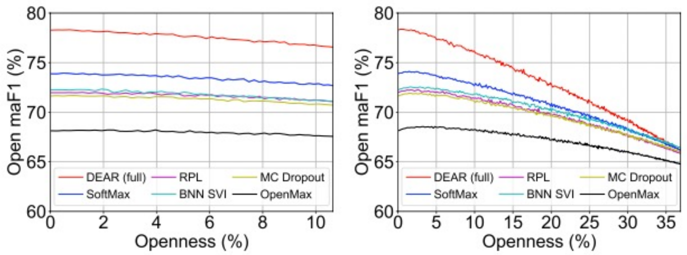
- Comparison with SoTA models

Table 1: **Comparison with state-of-the-art methods.** Models are trained on the closed set UCF-101 [55] and tested on two different open sets where the samples of unknown class are from HMDB-51 [31] and MiT-v2 [39], respectively. For Open maF1 scores, both the mean and standard deviation of 10 random trials of unknown class selection are reported. Closed set accuracy is for reference only.

Models	OSAR Methods	UCF-101 [55] + HMDB-51 [31]		UCF-101 [55] + MiT-v2 [39]		Closed Set Accuracy (%) **(For reference only)**
		Open maF1 (%)	Open Set AUC (%)	Open maF1 (%)	Open Set AUC (%)	
I3D [8]	OpenMax [5]	67.85 ± 0.12	74.34	66.22 ± 0.16	77.76	56.60
	MC Dropout	71.13 ± 0.15	75.07	68.11 ± 0.20	79.14	94.11
	BNN SVI [27]	71.57 ± 0.17	74.66	68.65 ± 0.21	79.50	93.89
	SoftMax	73.19 ± 0.17	75.68	68.84 ± 0.23	79.94	94.11
	RPL [10]	71.48 ± 0.15	75.20	68.11 ± 0.20	79.16	94.26
	DEAR (ours)	**77.24** ± 0.18	**77.08**	**69.98** ± 0.23	**81.54**	93.89
TSM [35]	OpenMax [5]	74.17 ± 0.17	77.07	**71.81** ± 0.20	83.05	65.48
	MC Dropout	71.52 ± 0.18	73.85	65.32 ± 0.25	78.35	95.06
	BNN SVI [27]	69.11 ± 0.16	73.42	64.28 ± 0.23	77.39	94.71
	SoftMax	78.27 ± 0.20	77.99	71.68 ± 0.27	82.38	95.03
	RPL [10]	69.34 ± 0.17	73.62	63.92 ± 0.25	77.28	95.59
	DEAR (ours)	**84.69** ± 0.20	**78.65**	70.15 ± 0.30	**83.92**	94.48
SlowFast [14]	OpenMax [5]	73.57 ± 0.10	78.76	72.48 ± 0.12	80.62	62.09
	MC Dropout	70.55 ± 0.14	75.41	67.53 ± 0.17	78.49	96.75
	BNN SVI [27]	69.19 ± 0.13	74.78	65.22 ± 0.21	77.39	96.43
	SoftMax	78.04 ± 0.16	79.16	74.42 ± 0.22	82.88	96.70
	RPL [10]	68.32 ± 0.13	74.23	66.33 ± 0.17	77.42	96.93
	DEAR (ours)	**85.48** ± 0.19	**82.94**	**77.28** ± 0.26	**86.99**	96.48
TPN [62]	OpenMax [5]	65.27 ± 0.09	74.12	64.80 ± 0.10	76.26	53.24
	MC Dropout	68.45 ± 0.12	74.13	65.77 ± 0.17	77.76	95.43
	BNN SVI [27]	63.81 ± 0.11	72.68	61.40 ± 0.15	75.32	94.61
	SoftMax	76.23 ± 0.14	77.97	70.82 ± 0.21	81.35	95.51
	RPL [10]	70.31 ± 0.13	75.32	66.21 ± 0.21	78.21	95.48
	DEAR (ours)	**81.79** ± 0.15	**79.23**	**71.18** ± 0.23	**81.80**	96.30

Experiments

- Open maF1 scores
  - Dear가 어떠한 openess에서도 가장 좋은 성능을 보이고 있다



(a) HMDB-51 as Unknown (b) MiT-v2 as Unknown

Figure 6: **Open macro-F1 scores against varying Openness.** The maximum openness is determined by the number of unknown classes, i.e., in $\omega_O^{(i)}$, $i=51$ for HMDB-51 and $i=305$ for MiT-v2.

- Ablation study
  - TPN을 AR 모델로 사용
  - HMDB-51을 unknown으로 사용
  - Joint는 CED 학습을 의미

Table 2: **Ablation studies.** Based on TPN [62] model, HMDB-51 [31] is used as the unknown. Best results are shown in bold.

$\mathcal{L}_{EUC}$	CED	Joint Train	Open maF1 (%)	OS-AUC (%)
✗	✗	✓	74.95 ± 0.18	77.12
✓	✗	✓	75.88 ± 0.16	77.49
✓	✓	✗	81.18 ± 0.15	79.02
✓	✓	✓	**81.79 ± 0.15**	**79.23**

Experiments

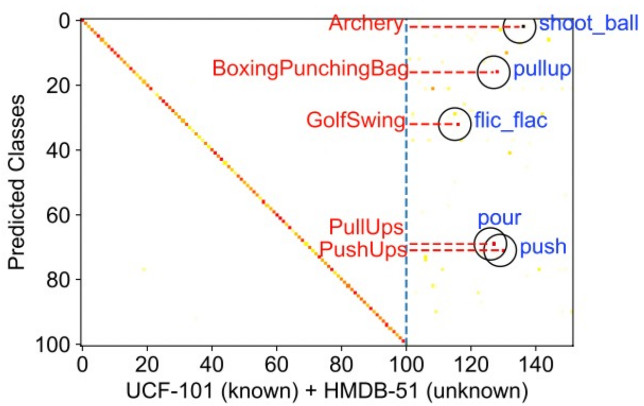
- Representation Debiasing
  - TSM model을 활용
  - Kinetics와 Mimetics dataset을 이용해서 debiasing 효과를 확인

Table 4: **Accuracy (%) on Biased and Unbiased dataset.**

Methods	Biased (Kinetics)		Unbiased (Mimetics)	
	top-1	top-5	top-1	top-5
DEAR (w/o CED)	91.18	99.30	26.56	69.53
DEAR (full)	**91.18**	**99.54**	**34.38**	**75.00**

Figure 21: **Examples of Kinetics and Mimetics.** The check mark (✓) indicates that the predicted label is correct while the cross mark (✗) means that the predicted label is incorrect.

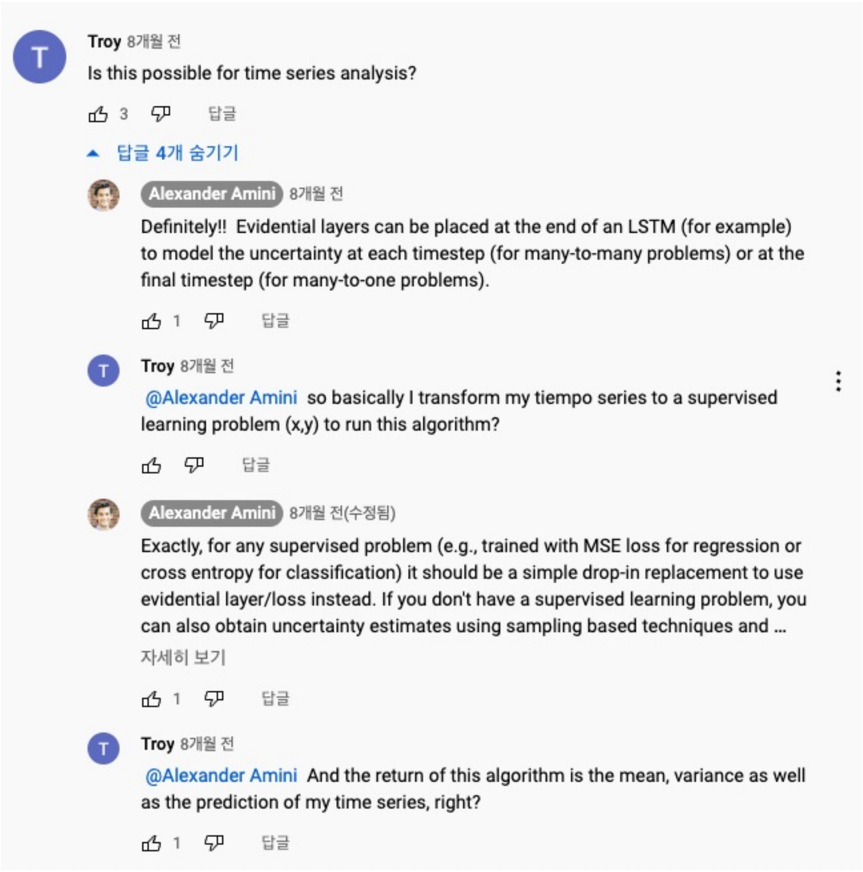
- 어떤 class를 많이 틀렸을까?
  - Shootball (unknown) → Archery
  - Pullup (unknown) → BoxingPunchingBag
  - 분석을 통해 'background scene의 영향을 많이 받는다'라는 사실 확인 (static bias)



Application on our research?

- Weakly supervised가 아닌 Fully supervised 상황에서, open set training 가능
- Un(Weakly)supervised setting의 경우 closed set 학습 자체도 어렵기 때문에 아직은 적용에 어려움 있음
- Evidential Deep Learning의 경우에도 결국엔, NLL을 이용하기 때문에 정답 어느정도 정확도를 갖는 ground truth label이 필요하다고 생각됨
- Fully-supervised temporal action localization task에서는 활용해볼 수 있는 여지가 있다고 판단

- From Youtube comments of the author’s lecture at MIT
  - Time series analysis에서의 anomaly-detection 등에 활용 가능

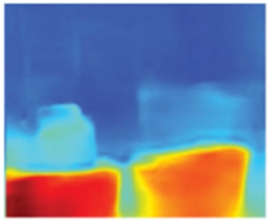


Application on other fields

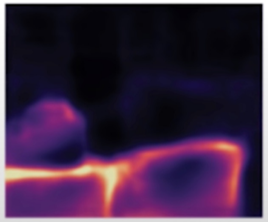
Applications of evidential learning

Monocular Depth Estimation

RGB input

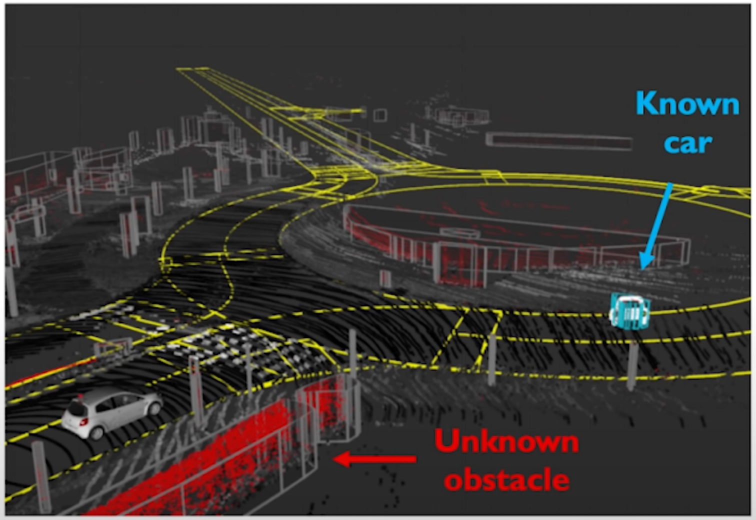


Predicted depth



Predicted uncertainty

LiDAR Object Classification



References

[1] W. J. Scheirer, A. de Rezende Rocha, A. Sapkota, and T. E. Boult, “Toward open set recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 7, pp. 1757–1772, Jul. 2013.

[2] Geng, C., Huang, S. J., & Chen, S. (2020). Recent advances in open set recognition: A survey. IEEE transactions on pattern analysis and machine intelligence, 43(10), 3614-3631.

[3] Amini, A., Schwarting, W., Soleimany, A., & Rus, D. (2020). Deep evidential regression. Advances in Neural Information Processing Systems, 33, 14927-14937.

[4] Bao, W., Yu, Q., & Kong, Y. (2021). Evidential deep learning for open set action recognition. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 13349-13358).

[5] Sensoy, M., Kaplan, L., & Kandemir, M. (2018). Evidential deep learning to quantify classification uncertainty. Advances in Neural Information Processing Systems, 31.

* What is Dirichlet

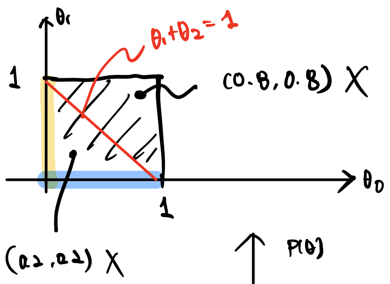
W. D state categorical R.V.
↳ $\theta \in \mathbb{R}^D$.

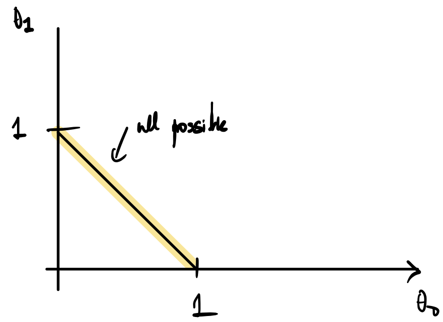
Constraint.

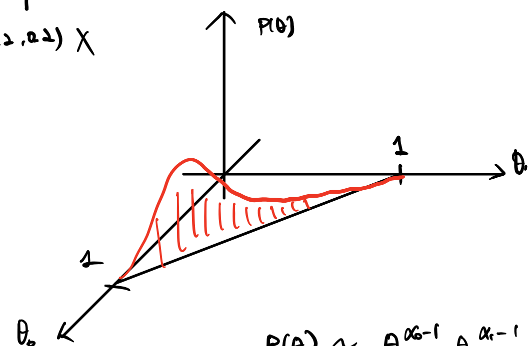
$$\begin{cases} \theta_d \in [0,1] \quad \forall d \\ \sum_{d=0}^{D-1} \theta_d = 1 \end{cases}$$

→ how to encode?

D=2 "





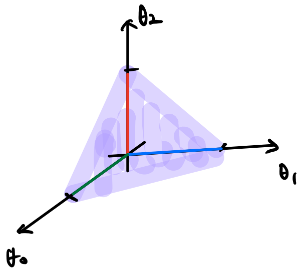


$$p(\theta) \sim \theta_0^{\alpha_0-1}\,\theta_1^{\alpha_1-1} \qquad \alpha_0, \alpha_1 > 0$$

Recall. β-dist

$p(\theta) \sim \theta^{\alpha-1}(1-\theta)^{\beta-1}$ → 2 state categorical

Say D=3



What is collection of point inside cube that meets the constraint

$$p(\theta) \sim \theta_0^{\alpha_0-1}\,\theta_1^{\alpha_1-1}\,\theta_2^{\alpha_2-1} \qquad \alpha_0, \alpha_1, \alpha_2 > 0$$

↗ extension of beta distribution.

θ는 D-1 dimension에 존재하게 된다.

$$p_k = b_k + a_k u, \quad \forall y \in \mathbb{Y} \tag{7}$$

$\sum_{k=1}^{K} p_k = \sum b_k + u = 1$ (8)

$$u + \sum_{k=1}^{K} b_k = 1 \tag{8}$$

$[p_1 \cdots p_K] \sim \mathrm{Dir}(p|\alpha)$

$$\mathrm{Dir}(\mathbf{p}|\boldsymbol{\alpha}) = \begin{cases} \frac{1}{B(\boldsymbol{\alpha})} \prod_{k=1}^{K} p_k^{\alpha_k - 1}, & \text{for } \mathbf{p} \in \mathcal{S}_K, \\ 0, & \text{otherwise,} \end{cases} \tag{9}$$

Dirichlet strength $S = \sum_{k=1}^{K} \alpha_k$

evidence $e = \{e_1 \cdots e_K\}$

by evidence theory

$$\boldsymbol{\alpha} = \mathbf{e} + \mathbf{a}W \tag{10}$$

$W = K \quad a_k = \frac{1}{K}$

$\alpha_k = e_k + 1$

$$\mathbb{E}(p_k) = \frac{\alpha_k}{\sum_{k=1}^{K} \alpha_k} = \frac{e_k + a_k W}{W + \sum_{k=1}^{K} e_k} \tag{11}$$

$\hat{p}_k = \frac{\alpha_k}{S}$

$p_k = b_k + \frac{1}{K} u$

이렇게 modeling 가능!

$$b_k = \frac{e_k}{S} \quad \text{and} \quad u = \frac{K}{S} \tag{12}$$

다시 식 (7)

$$p_k = b_k + a_k u, \quad \forall y \in \mathbb{Y} \tag{7}$$