Weekly Study Report - Evidential Deep Learning

Single Deterministic Model to Quantify Uncertainty

Wang Ma

2024-11-18

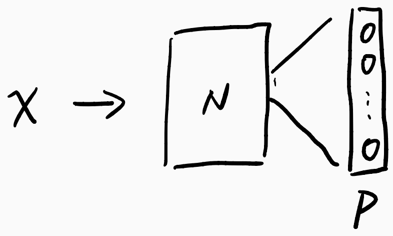
Electrical, Computer, and Systems Engineering Department
Rensselaer Polytechnic Institute

Outline

1. Foundations of Evidential Deep Learning

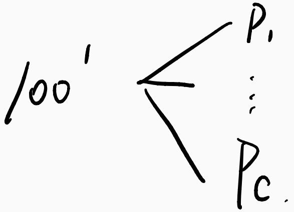
1.1 Evidential Deep Learning for Classification Uncertainty

Traditional NN:

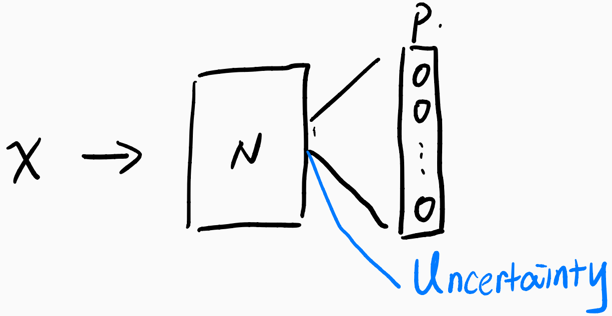


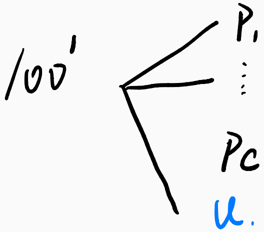
Model has 100 points.

⇒ distribute to C classes



Evidential NN:





Applied Dirichlet-distribution

$Dir(\alpha_1, \dots \alpha_c)$. $\alpha_i = e_i + 1$ logits

$u = \dfrac{C}{S}$, $\hat{p}_i = \dfrac{\alpha_i}{S}$, where $S = \sum \alpha_i$

1.2 Evidential Regression

Traditional NN:

$$x \to \boxed{\begin{matrix} NN \\ \Theta \end{matrix}} \to y$$

Evidential Regression.

$$y \sim N(\mu(x,\theta), \sigma^2(x,\theta))$$

$$y \sim N(\mu, \sigma^2)$$

$$\mu \sim N(\gamma, \sigma^2 \nu^{-1}), \quad \sigma^2 \sim \Gamma^{-1}(\alpha, \beta)$$

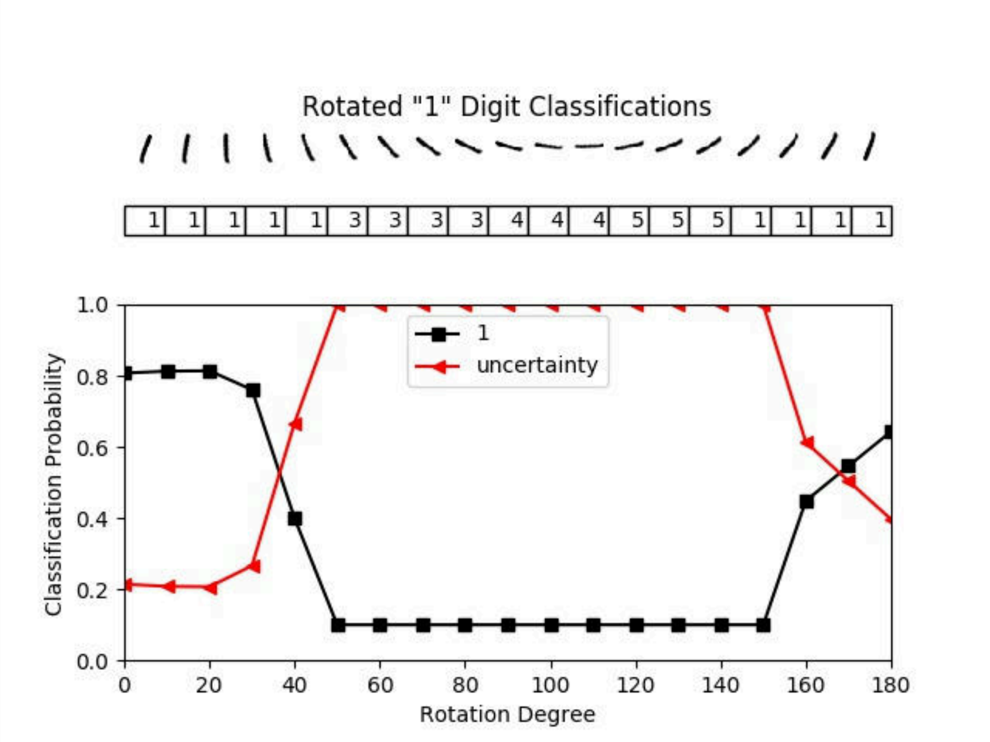
$$m = \{\gamma, \nu, \alpha, \beta\} \qquad \theta = \{\mu, \sigma^2\}$$

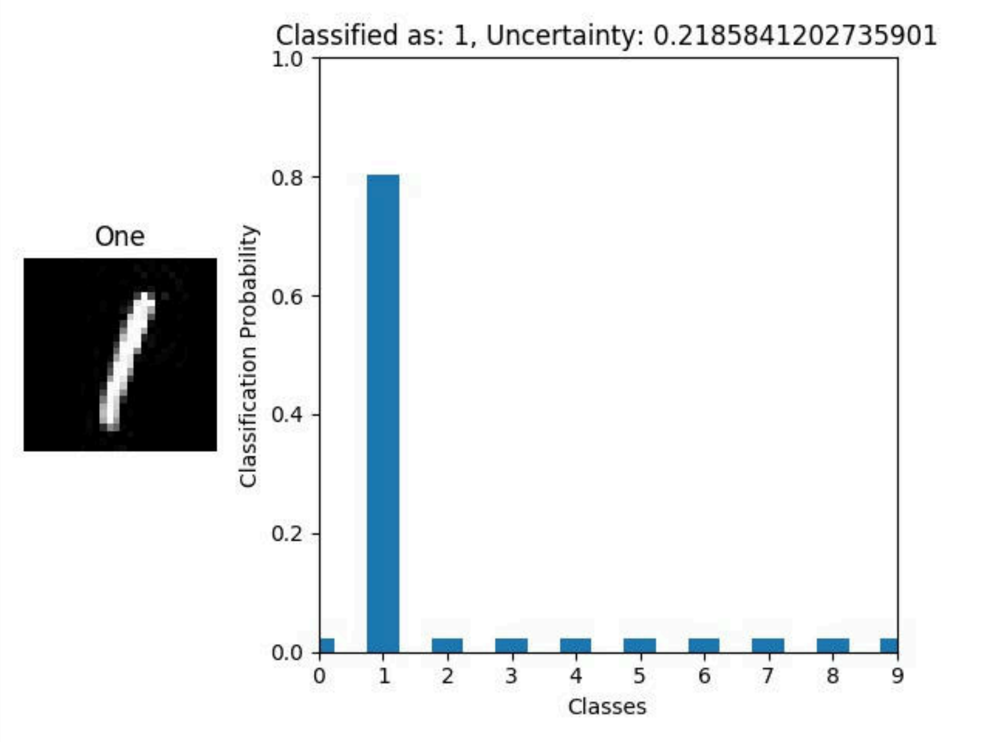
$$x \to \boxed{NN} \to \underbrace{\begin{Bmatrix} \gamma \\ \nu \\ \alpha \\ \beta \end{Bmatrix}}_{m}$$

$$\Rightarrow \underbrace{y = E(\mu) = \gamma}_{\text{prediction}}, \quad \underbrace{E(\sigma^2) = \frac{\beta}{\alpha - 1}}_{\text{Aleatoric}}, \quad \underbrace{Var(\mu) = \frac{\beta}{\gamma(\alpha-1)}}_{\text{Epistemic}}$$

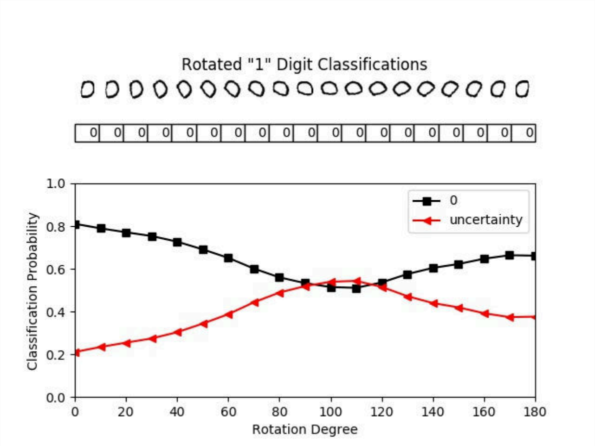
2. Experiments to Reproduce EDL

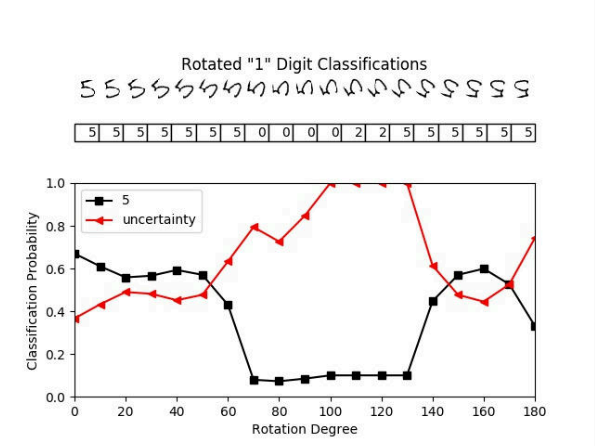
2.1 Reproduce Original Paper's Results

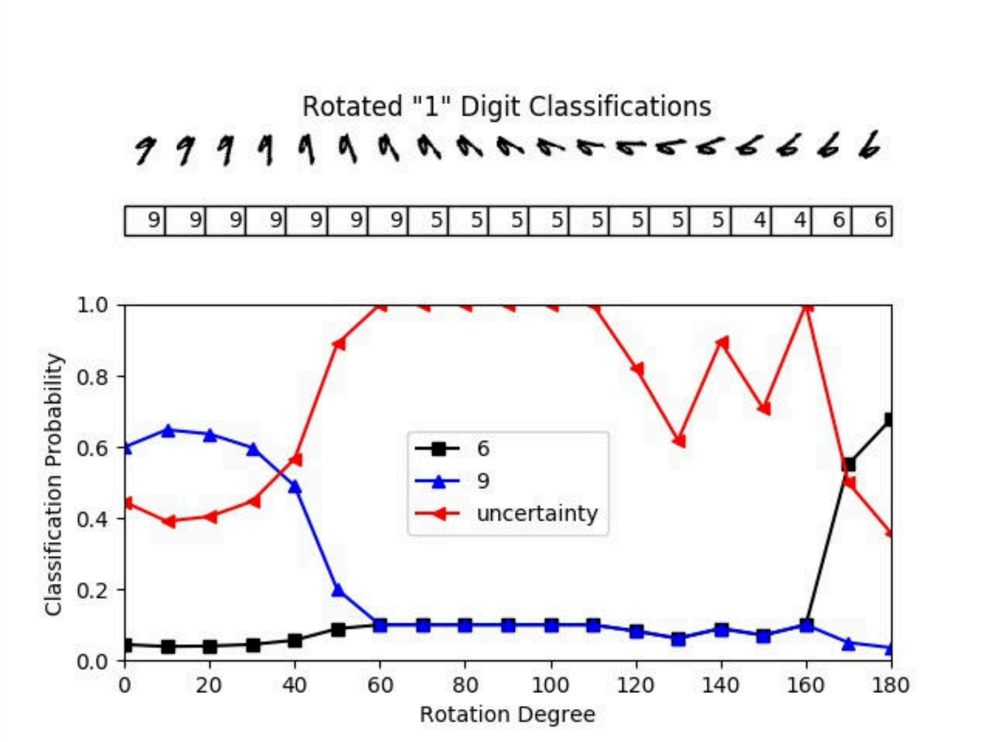


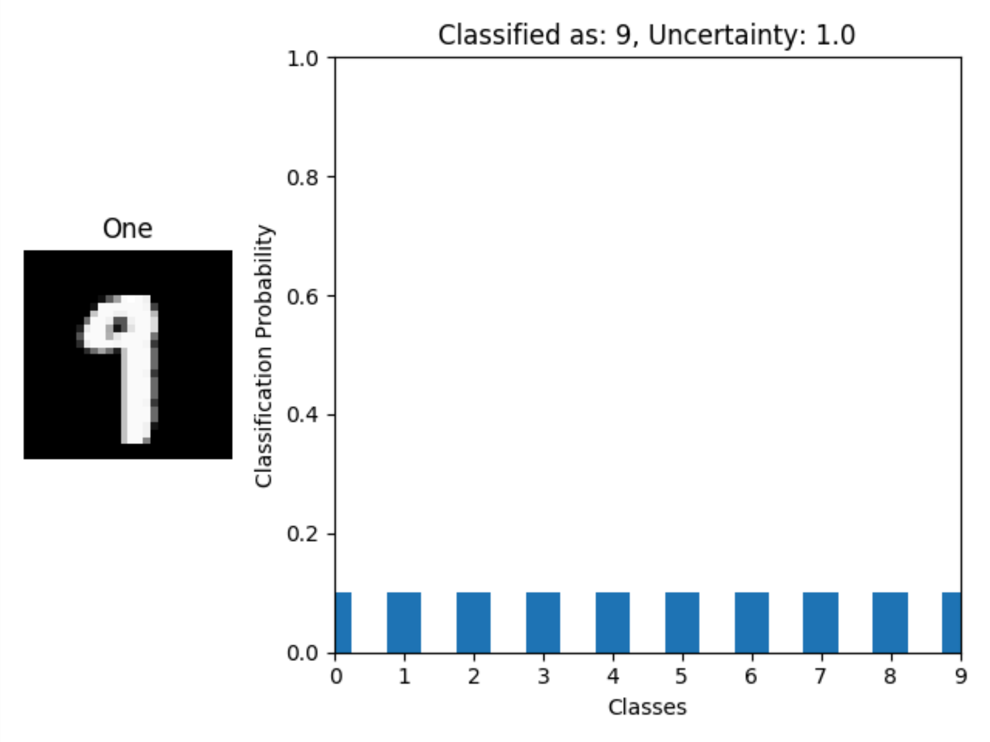


2.2 Using New Single Images to Test

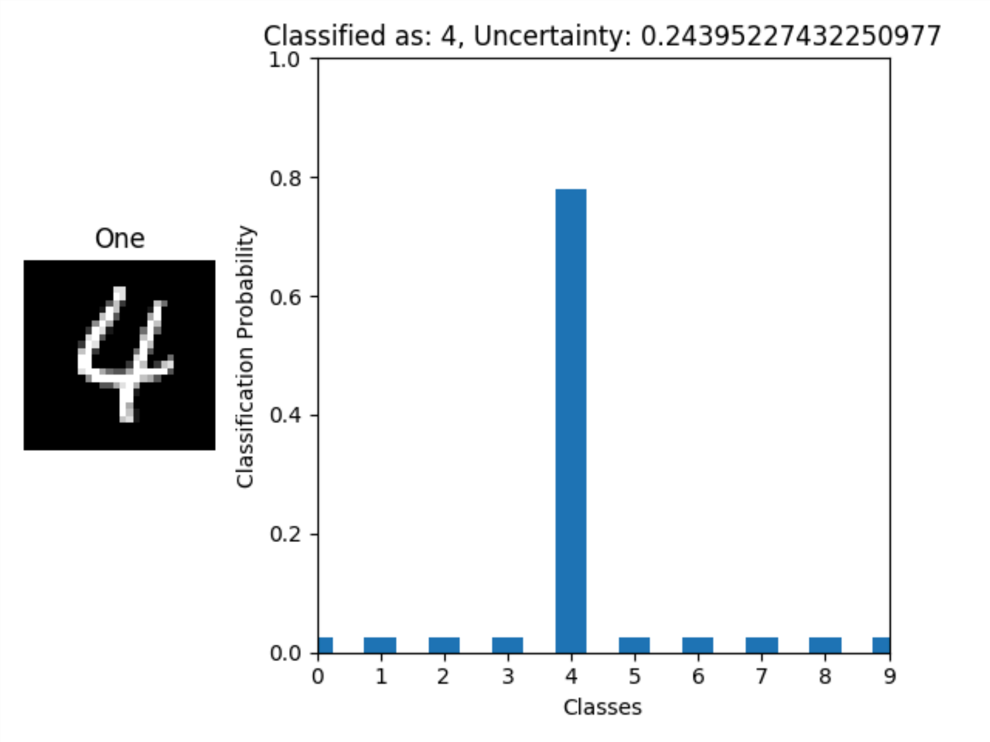








(Recognized High uncertainty by Ensemble)



(Recognized Low Uncertainty by Ensemble)

3. To Measure Label-~~wise~~ Aware Uncertainty

3.1 Hyper Evidential Deep Learning to Quantify Composite Classification Uncertainty (ICLR 2024)

Former:

$$\boxed{\text{100 points}} \Rightarrow \begin{cases} C_1 \\ C_2 \\ \vdots \\ C_K \\ u \end{cases}$$

K exclusive singleton sets.

Hyper-ENN

$$\boxed{\text{100 points}} \Rightarrow \begin{cases} C_1 \\ \vdots \\ C_{2^K-2} \\ u. \end{cases}$$

K labels $\begin{cases} L_1 \\ L_2 \\ \vdots \\ L_K \end{cases}$ $\Rightarrow$ 2^K-2 Label sets (no ϕ or full sets).

①	②
$\{L_1\}$	$\{L_1, L_2\}$
$\vdots$	$\{L_1, L_3\}$
$\{L_K\}$	$\{L_1, L_2, L_3\}$
	$\vdots$
single	composite.

3.2 Evidential Uncertainty Quantification: A Variance-Based Perspective (WACV 2024)

Treat $\underbrace{y \sim Cat(p)}_{r.v.}$, $p \sim Dir(\alpha)$

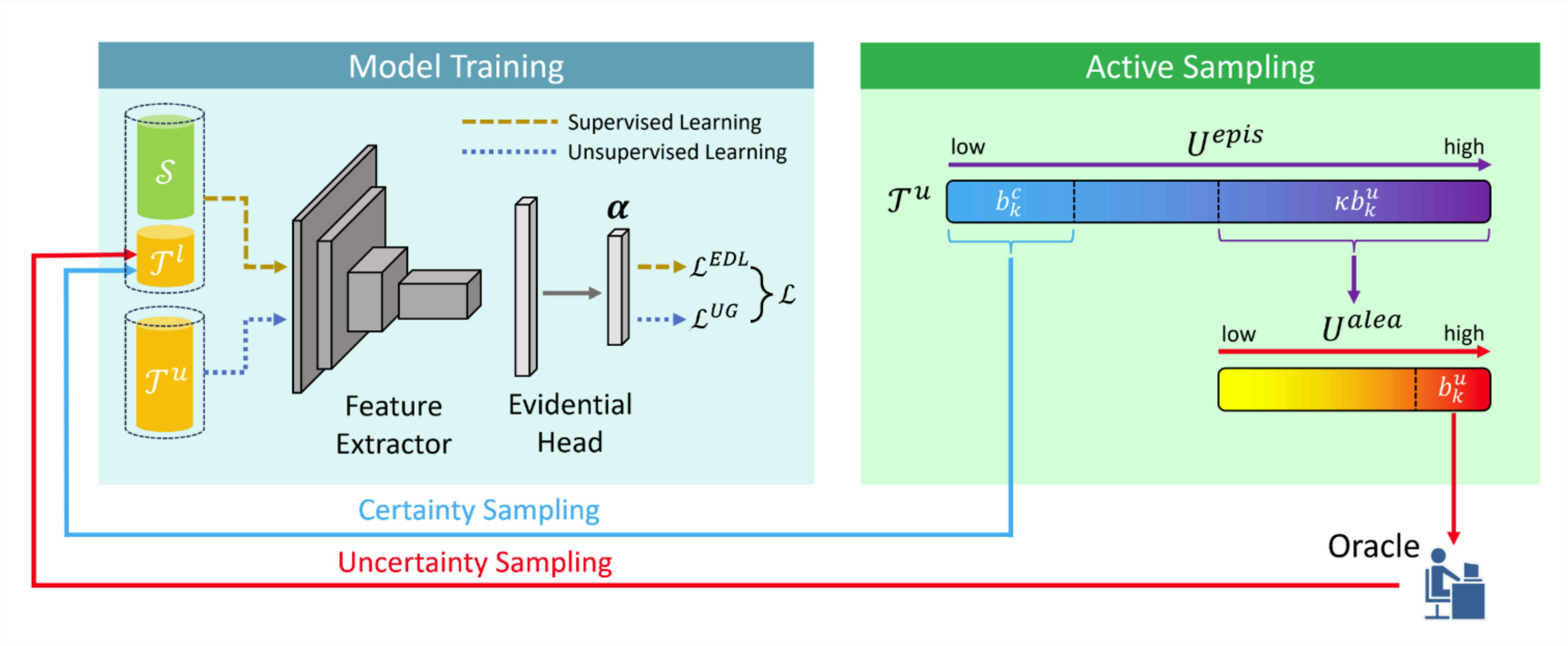
$$\underbrace{Cov[y]}_{\text{total}} = \underbrace{E[Cov(y|p)]}_{\text{Aleatoric}} + \underbrace{Cov[E(y|p)]}_{\text{epistemic.}}$$

M^T $\quad$ M^A $\quad$ M^E

$$M^T_{c,c} = \bar{p}_c (1 - \bar{p}_c)$$

$$M^A_{c,c} = \frac{\alpha_0}{\alpha_0 + 1} \bar{p}_c (1 - \bar{p}_c)$$

$$M^E_{c,c} = \frac{1}{\alpha_0 + 1} \bar{p}_c (1 - \bar{p}_c) . \qquad \alpha_0 = \sum \alpha_i$$


Model Training
Active Sampling
Supervised Learning
Unsupervised Learning
$\mathcal{S}$
$\mathcal{T}^l$
$\mathcal{T}^u$
α
$\mathcal{L}^{EDL}$
$\mathcal{L}^{UG}$
$\mathcal{L}$
Feature Extractor
Evidential Head
low
U^{epis}
high
b_k^c
κb_k^u
U^{alea}
b_k^u
Certainty Sampling
Uncertainty Sampling
Oracle

3.3 Conformal Prediction for Deep Classifier via Label Ranking (ICML 2024)

3.3 Conformal Prediction for Deep Classifier via Label Ranking (ICML 2024)

$$X \longrightarrow \boxed{NN} \rightarrow \{\_\,,\ \_\}$$

$\Uparrow$

To find the most composite set of possible labels.

① Model is uncertain

② which specific part is model uncertain?

4. Applications: EDL Enhanced Machine Learning

4.1 EDL Enhanced Multiple Instance Learning (MIL)

Multiple Instance Learning: A bag of Data $\{x_i\}^N$ have the label $y_B = \{0, 1\}^C$. If $y_B[j] = 1$, then at lease one of x_i belongs to jth Class.

This task want to do: Bag-level prediction and Instance-level prediction. Obviously, EDL introduces the Uncertainty Quantification to MIL task. And more:

1. If we do Instance-level prediction, we have estimated α_i for every instance, so we have $p_i \sim Dir(\alpha_i)$, and so instance-level uncertainties.
2. If we do Bag-level prediction, we can dynamically adjust the evidence contribution from different instances.

$$e^{bag} = \sum_{j=1}^{N} w_j e_j^{instance}, \text{ where } w_j = \frac{1}{1 + Unc(x_j)}.$$

4.2 EDL Enhanced Transfer Learning

4.2.1 Domain Adaptation

1. **Labeling:**

$$y^t = \begin{cases} j & \text{if } logS^t \geq \delta_j, j = \arg\max_{1 \leq k \leq L_s} \alpha_k^t \\ unknown & \text{if } logS^t < \delta_j, j = \arg\max_{1 \leq k \leq L_s} \alpha_k^t \end{cases}$$

2. **Aligning Evidence Vector:**

- x_s from source domain, using a feature extractor φ and we get $f_s = \varphi(x_s)$;
- x_t from target domain, we have $f_t = \varphi(x_t)$.

Then let the EDL model be g, we have the evidence vector $m^s = g(f_s)$ and $m^t = g(f_t)$. Then we align these two vectors with criteria D:

$$L_{align} = \sum_{k=1}^{N_s} D(m_k^s) - \sum_{j=1}^{N_T} D(m_j^T).$$

4.2.2 Model Adaptation

1. Distillation (like Hanjing's Work): $L_{KD} = D(P_s \parallel P_T) + \lambda D(U_s \parallel U_T)$,

   where P_s and P_T are predictive distributions, U_s and U_T are predictive uncertainties.

2. Meta Learning approach:

$$\Theta_{final} = \Theta_{meta} + \Theta_{pretrained}$$

   .

   This is like a updating process from Prior to Posterior.

4.3 EDL Enhanced Active Learning

4.3 EDL Enhanced Active Learning

1. **Uncertainty-based Sampling**: weighted sampling strategies:

   $$\text{Uncertainty Score} = \omega_1 AU + \omega_2 EU.$$

2. **Second-order Uncertainty**: Vacuity and Dissonance.

   - Vacuity: the uncertainty due to a lack of evidence in the model's predictions
   - Dissonance: the conflict or inconsistency among the evidence gathered by the model for different classes

   Enhanced strategies:

   1. Select samples with High Vacuity to let model cover more area at first (more general)
   2. Select samples with High Dissonance to let model address the internal conflict (locally or specifically)

4.4 EDL Enhanced Multi-View Classification

An example:

A result from to views: $w^1 = \{b_i^1, u^1\}$ and $w^2 = \{b_i^2, u^2\}$, we have

$b_c = \frac{1}{1 - Conf}(b_i^1 b_i^2 + b_i^1 u^2 + b_i^2 u^1)$ and $u = \frac{1}{1 - Conf} u^1 u^2$

where $Conf = \sum_{i=1}^{C} b_i^1 b_i^2$.

Problem.

- This methods only consider the consistency among different views, while do not consider the complementarity ;
- When b_i are uniformly about 0, the Conf might do not work.

4.5 EDL Enhanced Multi-Label Learning

4.5 EDL Enhanced Multi-Label Learning

Core idea. For every label, the EDL output a $\{\alpha_i, \beta_i\}$ for a Beta Distribution, as the evidence for label i (based on the input x).

So for every input x, we have its prediction for every label i: $p_i \sim Beta(\alpha_i, \beta_i)$.

Then the predictive probablity:

$$p_i = \frac{\alpha_i}{\alpha_i + \beta_i},$$

and uncertainty:

$$u_i = H[Beta(\alpha_i, \beta_i)].$$

The problem: imbalance among labels and the ignorance of dependences among labels.