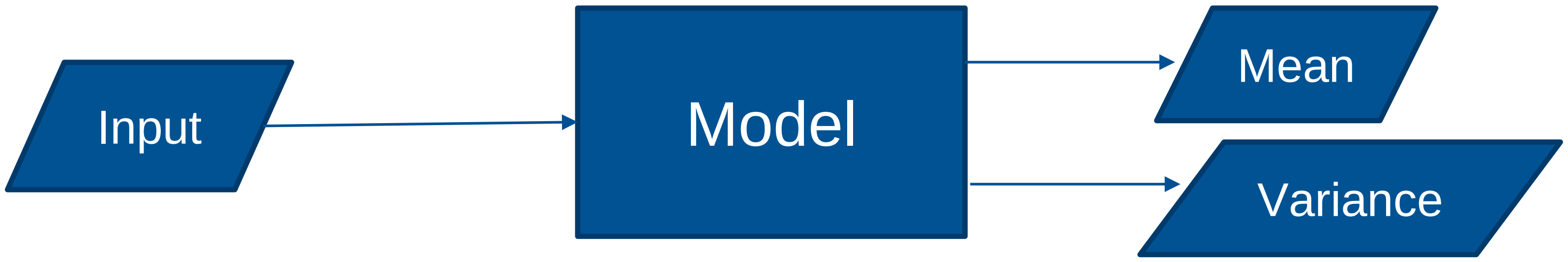
Special Topics in Applications (AIL861)
Artificial Intelligence for Earth Observation
Lecture 15

Instructor: Sudipan Saha

Ensemble – Look Back



Strategy: Sampling an ensemble of models to get different mean, variance values.

Evidential Learning

$y \in \{1, 2, \ldots, K\}$

$y \sim \mathrm{Categorical}\ (p)$

Class labels

Probabilities or distribution parameters

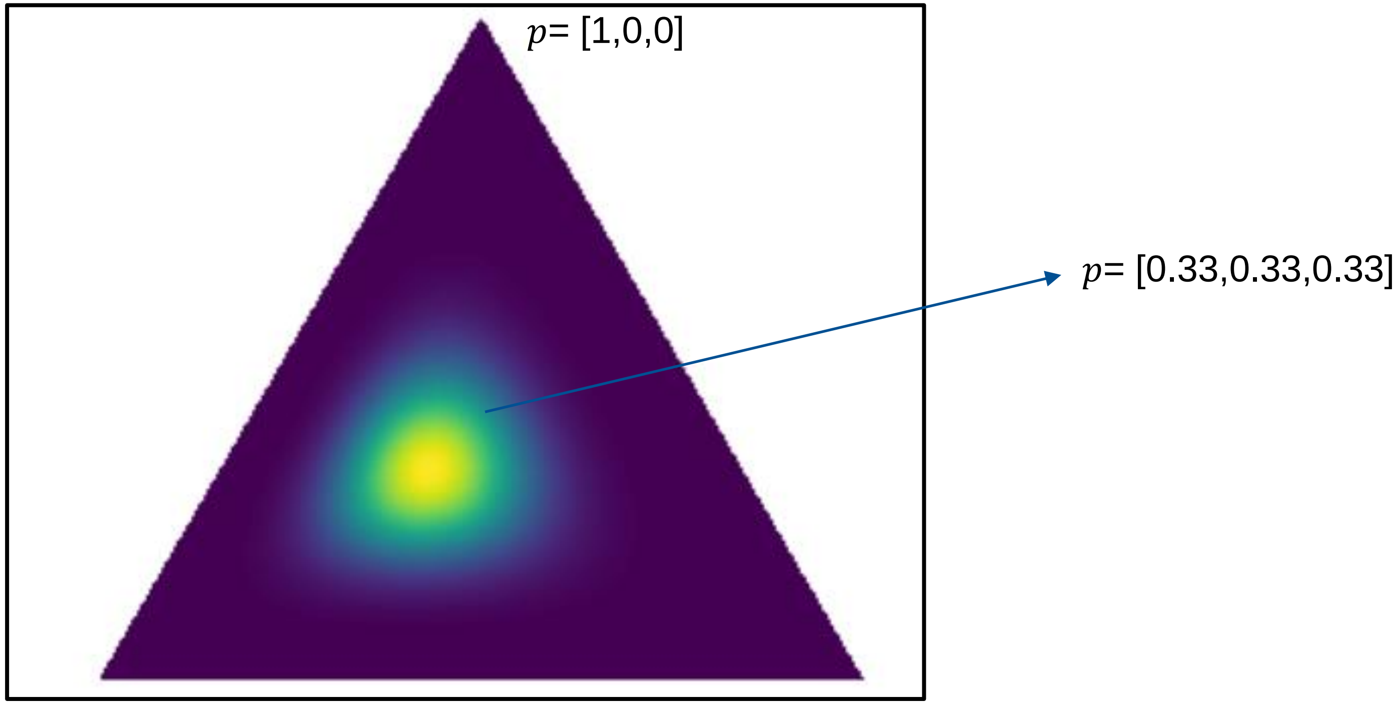
$p \sim Dirichlet\ (\alpha)$

Model parameters

Choice of evidential distribution is related to conjugate priors in Bayesian inference.

Sampling from Evidential Distribution

Sampling from evidential distribution yields individual new distribution over the data.



Dirichlet([10., 7., 8.]), K = 3

Dirichlet Distribution

- The Dirichlet distribution can be interpreted as a distribution over categorical distributions.

- While probability vector given by a softmax represents a single point on the underlying solution simplex, the Dirichlet distribution represents a distribution on this simplex.

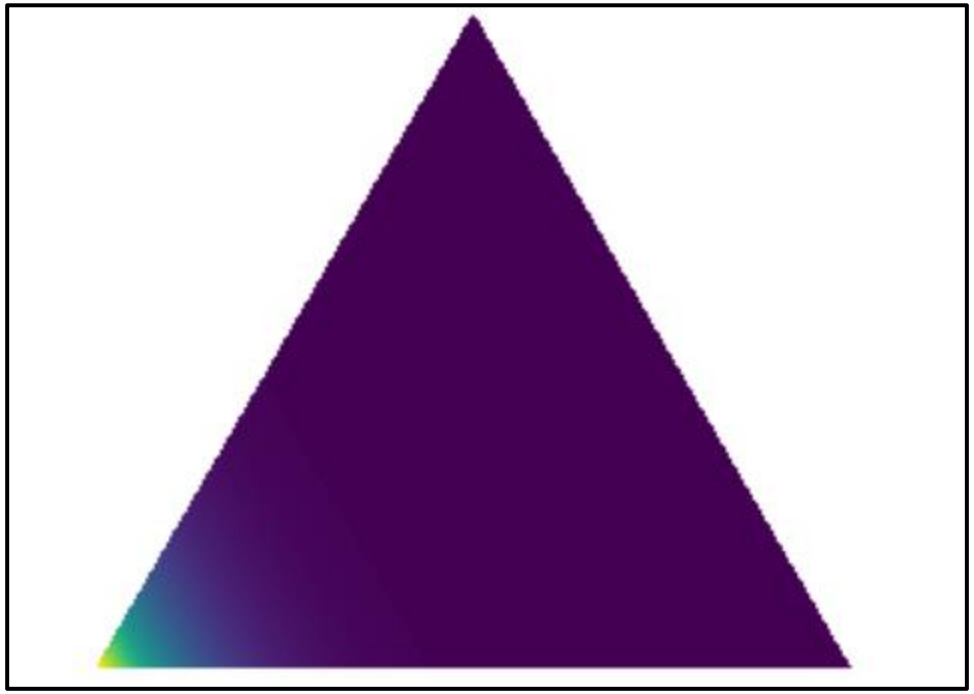
Dirichlet Prior

- ✓ Idea is to use the logits (z_c for class c) of the network to define the parameters of the Dirichlet distribution
- ✓ Parameter α_c is defined as $\exp(z_c)$

Different Input Scenarios

In domain confident: high class concentration for a single class

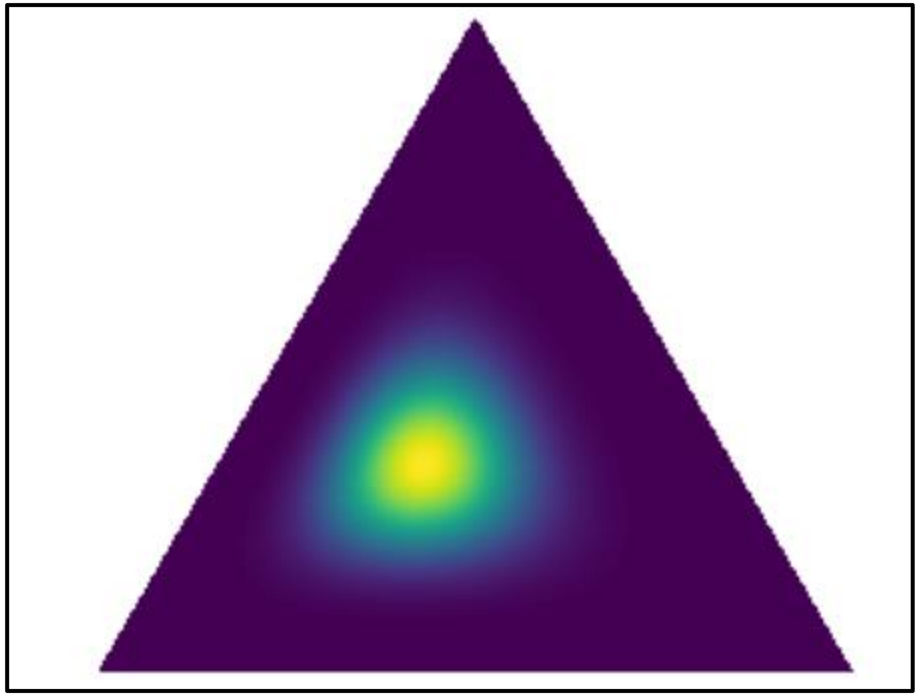
Dirichlet([10., 1., 1.])



Different Input Scenarios

Aleatoric uncertainty – class overlap: high class concentration for multiple or all classes

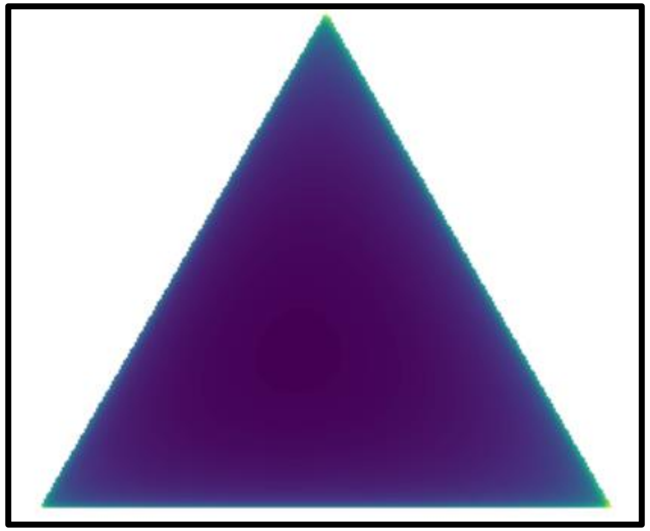
Dirichlet([10., 7., 8.])



Different Input Scenarios

OOD: low class concentration for all classes.

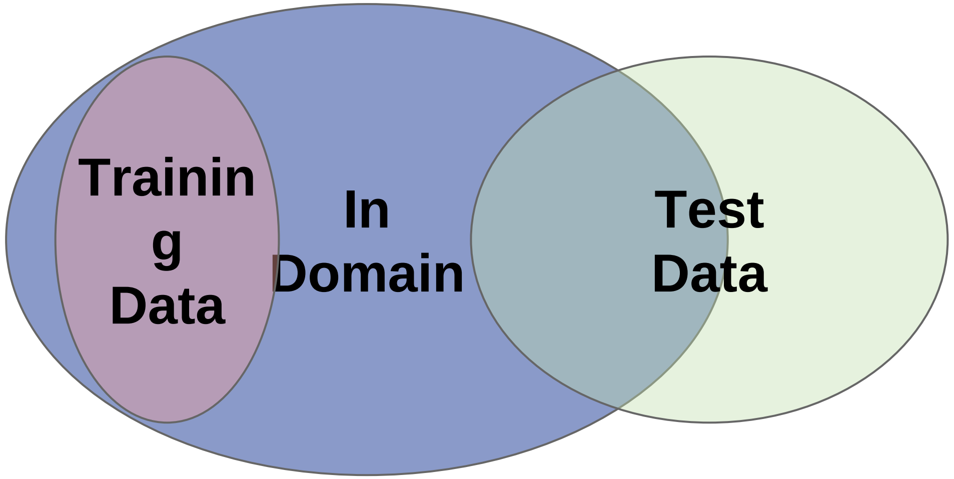
Dirichlet([0.94, 0.955, 0.952]



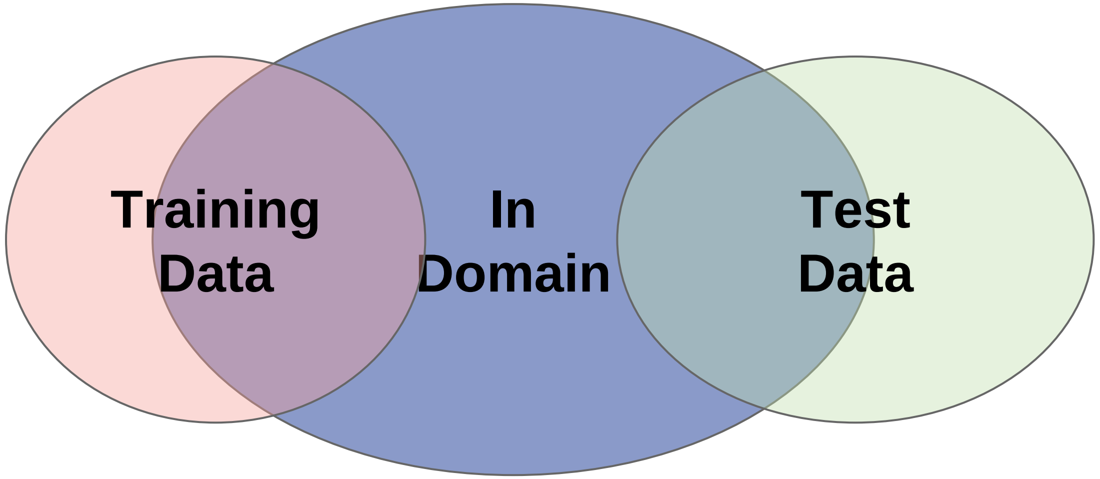
exp(-0.026) = 0.94

In other words, we want negative logits for the OOD samples

Unsupervised OOD



Supervised OOD



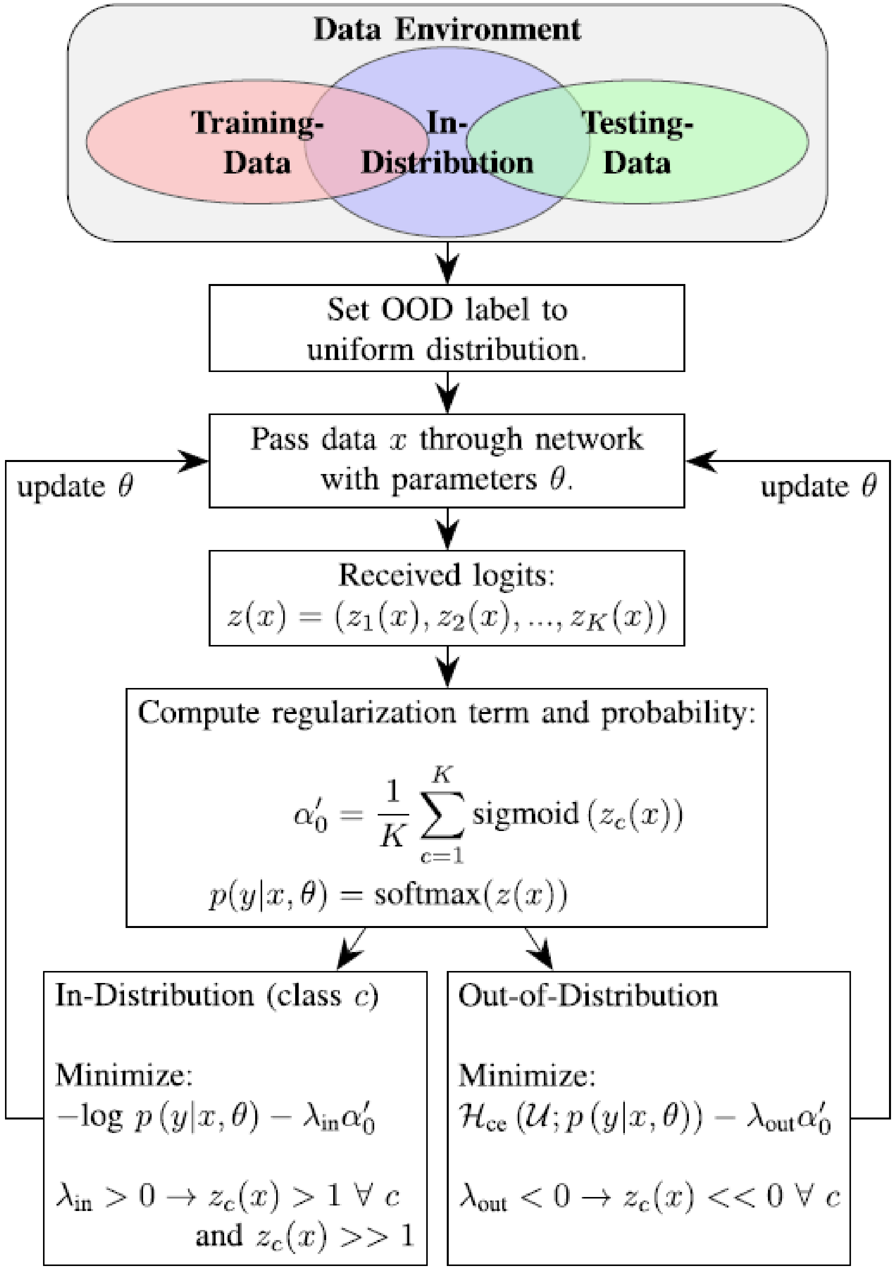
Loss Function

✓ For in-domain samples, use cross-entropy loss as usual

$$\mathcal{L}_{\text{in}}(\theta; \lambda_{\text{in}}) := \mathbb{E}_{P_{\text{in}}(x,y)}\left[-\log p(y|x,\theta) - \lambda_{\text{in}}\alpha'_0\right]$$

✓ For the OOD samples, use cross entropy loss computed against a uniform distribution over all classes

$$\mathcal{L}_{\text{out}}(\theta; \lambda_{\text{out}}) := \mathbb{E}_{P_{\text{out}}(x,y)}\left[\mathcal{H}_{\text{ce}}(\mathcal{U};\, p(y|x,\theta)) - \lambda_{\text{out}}\alpha'_0\right]$$


Data Environment
Training-Data
In-Distribution
Testing-Data
Set OOD label to uniform distribution.
Pass data x through network with parameters θ.
update θ
update θ
Received logits:
$z(x) = (z_1(x), z_2(x), ..., z_K(x))$
Compute regularization term and probability:
$\alpha_0' = \frac{1}{K} \sum_{c=1}^{K} \text{sigmoid}\left(z_c(x)\right)$
$p(y|x,\theta) = \text{softmax}(z(x))$
In-Distribution (class c)
Minimize:
$-\log\, p\left(y|x,\theta\right) - \lambda_{\text{in}}\alpha_0'$
$\lambda_{\text{in}} > 0 \to z_c(x) > 1\ \forall\ c$ and $z_c(x) >> 1$
Out-of-Distribution
Minimize:
$\mathcal{H}_{\text{ce}}\left(\mathcal{U}; p\left(y|x,\theta\right)\right) - \lambda_{\text{out}}\alpha_0'$
$\lambda_{\text{out}} < 0 \to z_c(x) << 0\ \forall\ c$

Experiments

Dataset

So2Sat LCZ42 Dataset

Contains Sentinel-1 and Sentinel-2 patches.

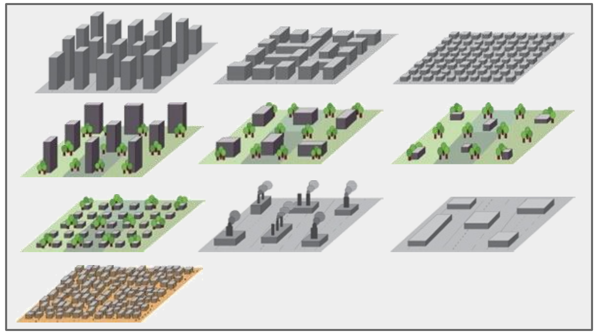
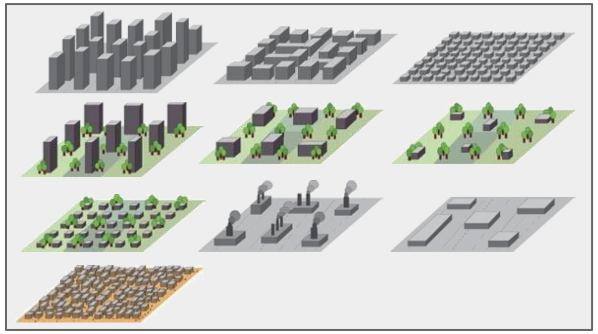
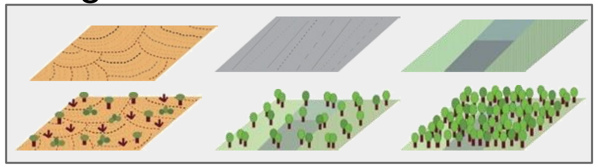
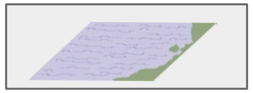
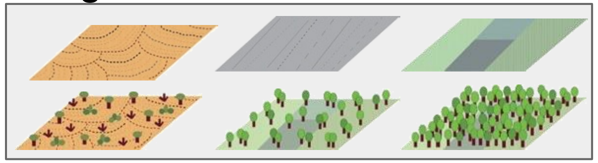
About half a million co-registered patches from 42 different regions.

Patches labeled as one of 17 local climate zones.

Our experiments are based on the Sentinel-2 data only.

AI4EO

Experiments

	Experiment 1	Experiment 2	Experiment 3
In-Domain	Urban 	Urban 	Red (Band B4)
OOD (Training)	Vegetation 	Water 	Green (Band B3)
OOD (Testing)	Water 	Vegetation 	Blue (Band B2)

AI4EO

Experiments

[redacted]

		Max. Prob.	Mutual Info.	α_0
Test Case 1	DPN^-	98.58 ± 0.89	99.35 ± 0.29	99.34 ± 0.29
	DPN_{forw}	95.87 ± 2.28	54.74 ± 9.97	50.59 ± 10.48
	ENN	75.64 ± 5.70	75.33 ± 4.46	76.75 ± 2.84
Test Case 2	DPN^-	78.65 ± 0.61	89.53 ± 0.53	89.67 ± 0.54
	DPN_{forw}	44.65 ± 5.09	34.21 ± 7.84	33.09 ± 7.47
	ENN	71.76 ± 0.90	71.75 ± 0.35	68.80 ± 2.23
Test Case 3	DPN^-	91.79 ± 0.20	95.52 ± 0.29	95.52 ± 0.38
	DPN_{forw}	71.89 ± 4.53	12.26 ± 3.10	11.80 ± 2.71
	ENN	58.89 ± 0.70	58.17 ± 1.23	56.83 ± 2.02

An Advanced Dirichlet Prior Network for Out-of-Distribution Detection in Remote Sensing

Jakob Gawlikowski, *Student Member, IEEE*, Sudipan Saha, *Member, IEEE*, Anna Kruspe, and Xiao Xiang Zhu, *Fellow, IEEE*

Observations: Open Set Recognition

- ✓ Is easy for high spatial resolution datasets.
- ✓ Is challenging for low spatial resolution datasets.

Observations: Sensor Shift

Contrary to the open-set recognition, the results indicate that the OOD detection under sensor shift is easier with lower resolution images and more challenging with higher resolution images. Furthermore, the similarity of the different sensors highly affects the OOD detection performance. It can be clearly observed that separating the blue channel from the green and the red channel gives the best results.

Observations: Region Shift

A region-wise shift is almost similarly prevalent in both urban classes and vegetation classes.