PyTorch Active Learning Library

Robert Munro, Human-in-the-Loop Machine Learning

An Active Learning library in Pytorch, implementing major techniques to sample items for annotation based on: model's confusion; gaps in the model's knowledge; and distributional properties of the data and target domain.

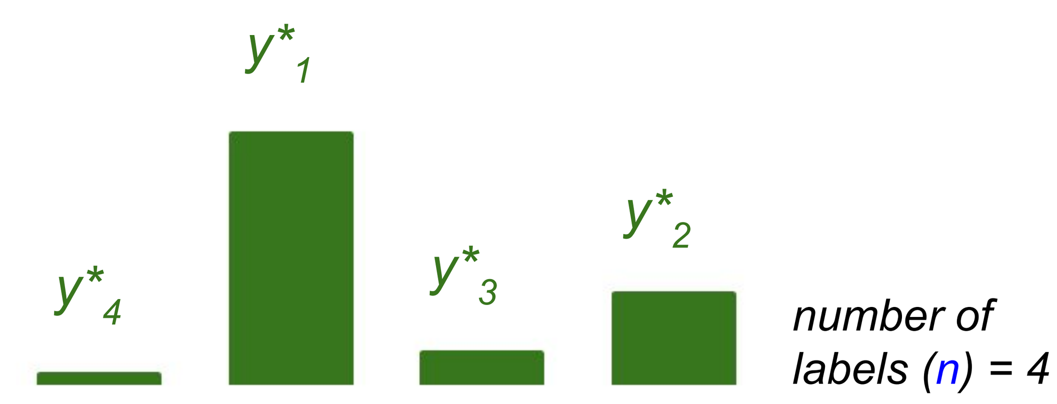
Machine Learning Knowledge Quadrant:

	Knowns	Unknowns
Known	Confident Model Predictions (Known Knowns) **Current Model State**	Non-Confident Model Predictions (Known Unknowns) **Uncertainty Sampling**
Unknown	Latent Information in Related Models (Unknown Knowns) **Transfer Learning**	Gaps in Model Knowledge (Unknown Unknowns) **Diversity Sampling**

Examples:

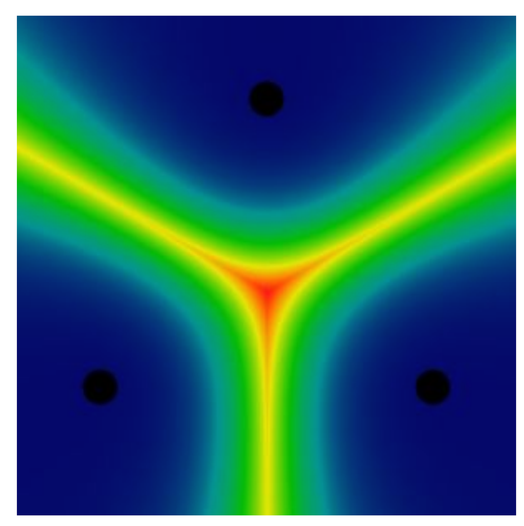
The heatmaps show the differences between the Uncertainty Sampling methods on a three-label dataset. The methods use a probability distribution like:

```
prob = torch.tensor([0.032, 0.643, 0.087, 0.236])
```



Least Confidence: *difference between the most confident prediction and 100% confidence*

$$\frac{n\,(1 - P_\theta(y^*_1 \mid x))}{n - 1}$$

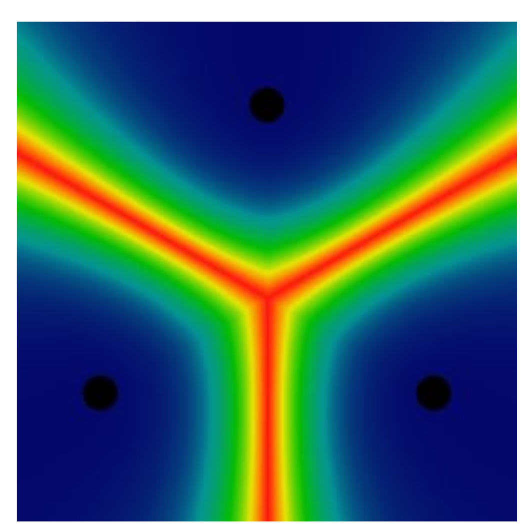


```
most_conf = torch.max(prob)
labs = prob.numel()
numerator = (labs * (1 - most_conf))
denominator = (labs - 1)

least_conf = numerator / denominator
```

Margin of Confidence: difference between the top two most confident predictions

$$1 - (P_\theta(y^*_1 \mid x) - P_\theta(y^*_2 \mid x))$$

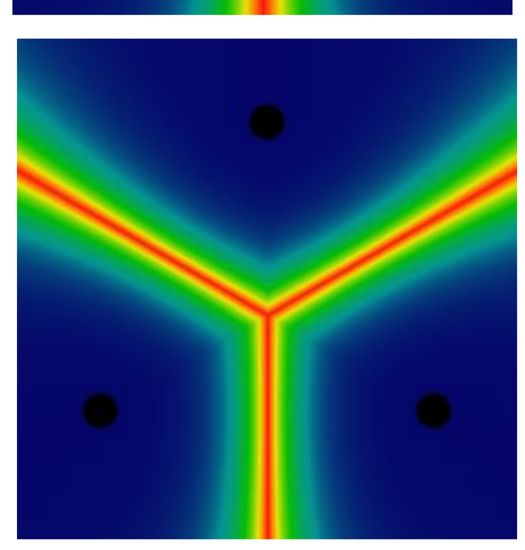


```
prob, _ = torch.sort(prob, descending=True)
difference = (prob.data[0] - prob.data[1])

margin_conf = 1 - difference
```

Ratio of Confidence: ratio between the top two most confident predictions

$$\frac{P_\theta(y^*_2 \mid x)}{P_\theta(y^*_1 \mid x)}$$

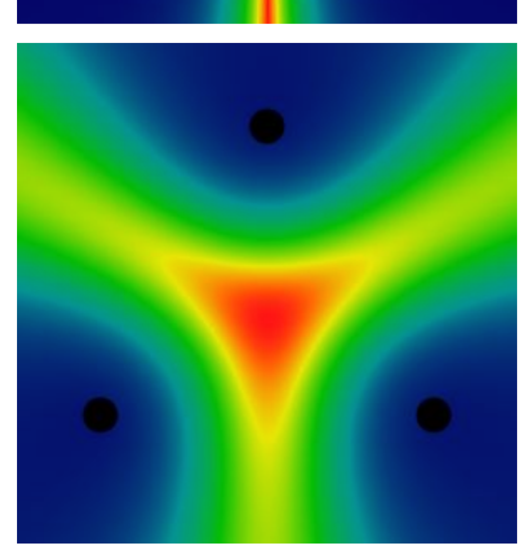


```
prob, _ = torch.sort(prob, descending=True)

ratio_conf = (prob.data[1] / prob.data[0])
```

Entropy: the difference between all predictions, as defined by information theory

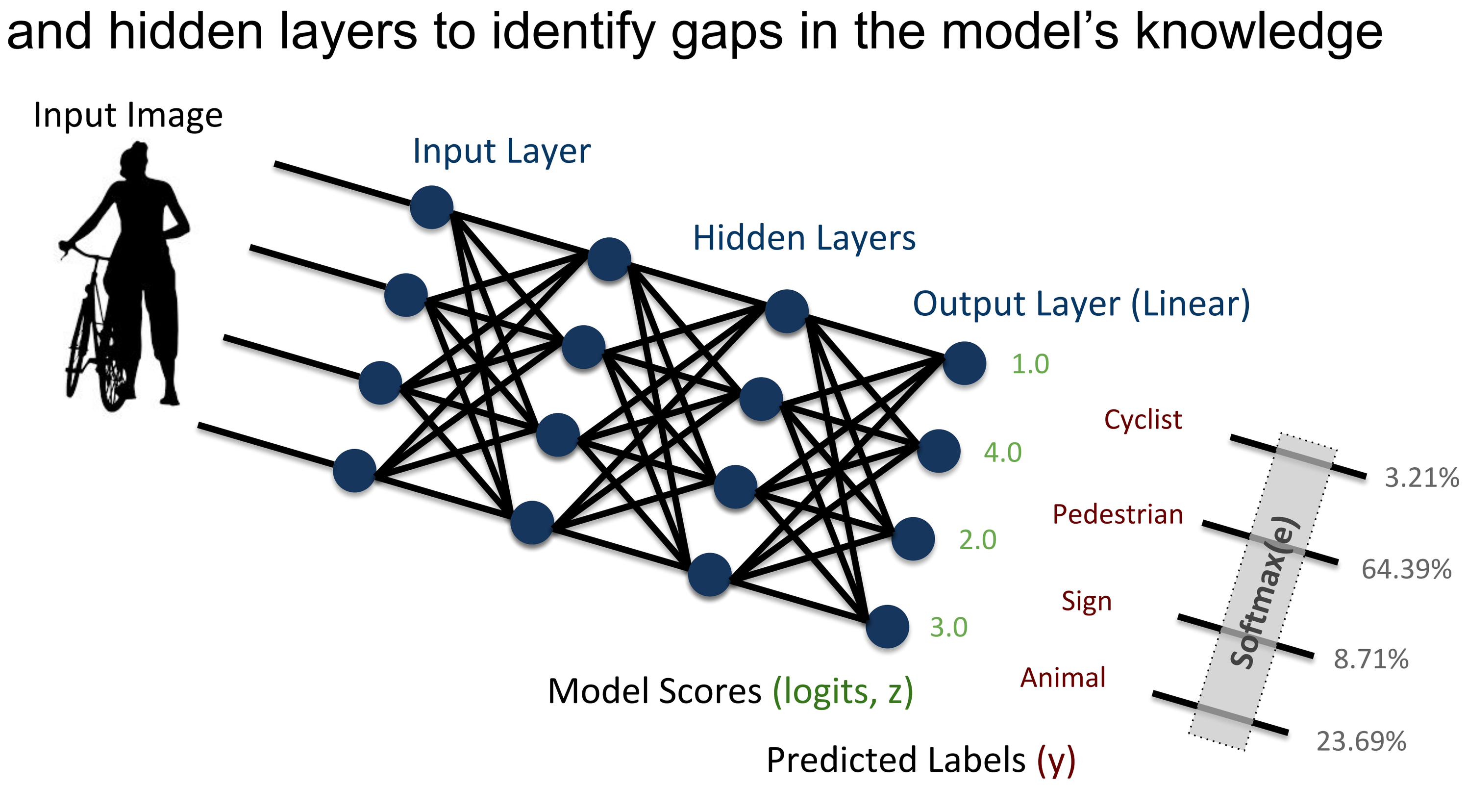
$$\frac{-\sum_y P_\theta(y \mid x)\,\log_2 P_\theta(y \mid x))}{\log_2(n)}$$



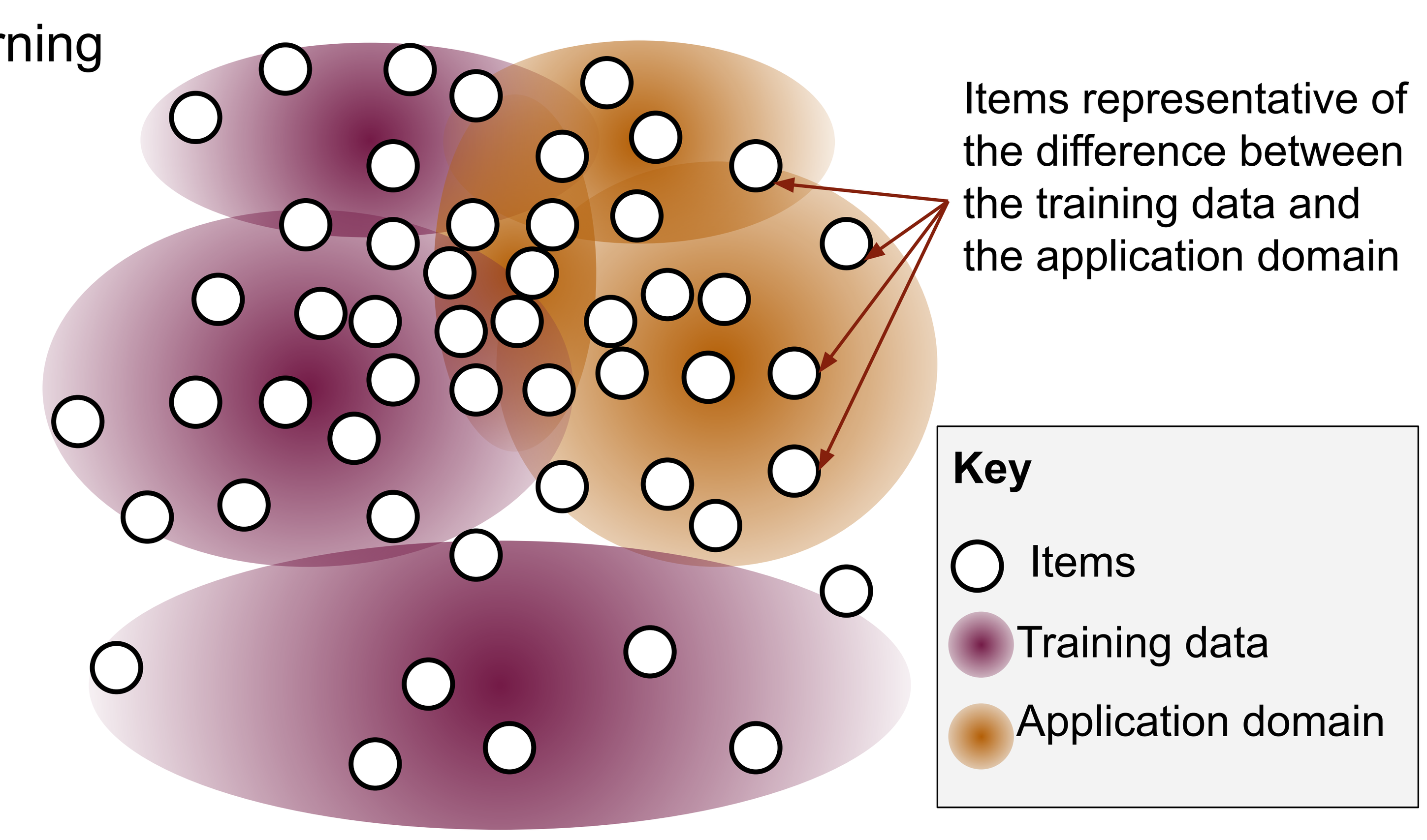
```
prbslogs = prob * torch.log2(prob)
numerator = 0 - np.sum(prbslogs)
denominator = math.log2(prob.numel())

entropy = numerator / denominator
```

Model-based Outliers: sampling for low activation in logits and hidden layers to identify gaps in the model's knowledge



Advanced Active Learning: combining multiple Active Learning techniques and incorporating Unsupervised Machine Learning, Domain Adaptation, and Transfer Learning



Code is open source & written to accompany:
Human-in-the-Loop Machine Learning, Robert Munro, *Manning Publications*. http://bit.ly/huml_book