Uncertainty and Introduction to Active Learning

Swati Mishra

Applications of Machine Learning (4AL3)

Fall 2024

McMaster University

ENGINEERING

Review

- Human Computation Architectures
- Crowdsourcing
- Computer Vision and Machine Learning Datasets
- Designing Annotation Interfaces

The concept of uncertainty

- What do probability distributions tell us?

	mean radius	mean texture	mean perimeter	mean area	mean smoothness	mean compactness	mean concavity	mean concave points	mean symmetry	mean fractal dimension	...	worst radius	worst texture	worst perimeter	
0	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.30010	0.14710	0.2419	0.07871	...	25.380	17.33	184.60	
1	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.08690	0.07017	0.1812	0.05667	...	24.990	23.41	158.80	
2	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.19740	0.12790	0.2069	0.05999	...	23.570	25.53	152.50	P_B = 0.99, P_M=0.01
3	11.42	20.38	77.58	386.1	0.14250	0.28390	0.24140	0.10520	0.2597	0.09744	...	14.910	26.50	98.87	
4	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.19800	0.10430	0.1809	0.05883	...	22.540	16.67	152.20	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
564	21.56	22.39	142.00	1479.0	0.11100	0.11590	0.24390	0.13890	0.1726	0.05623	...	25.450	26.40	166.10	P_B = 0.63, P_M=0.37
565	20.13	28.25	131.20	1261.0	0.09780	0.10340	0.14400	0.09791	0.1752	0.05533	...	23.690	38.25	155.00	
566	16.60	28.08	108.30	858.1	0.08455	0.10230	0.09251	0.05302	0.1590	0.05648	...	18.980	34.12	126.70	
567	20.60	29.33	140.10	1265.0	0.11780	0.27700	0.35140	0.15200	0.2397	0.07016	...	25.740	39.42	184.60	P_B = 0.12, P_M=0.88
568	7.76	24.54	47.92	181.0	0.05263	0.04362	0.00000	0.00000	0.1587	0.05884	...	9.456	30.37	59.16	

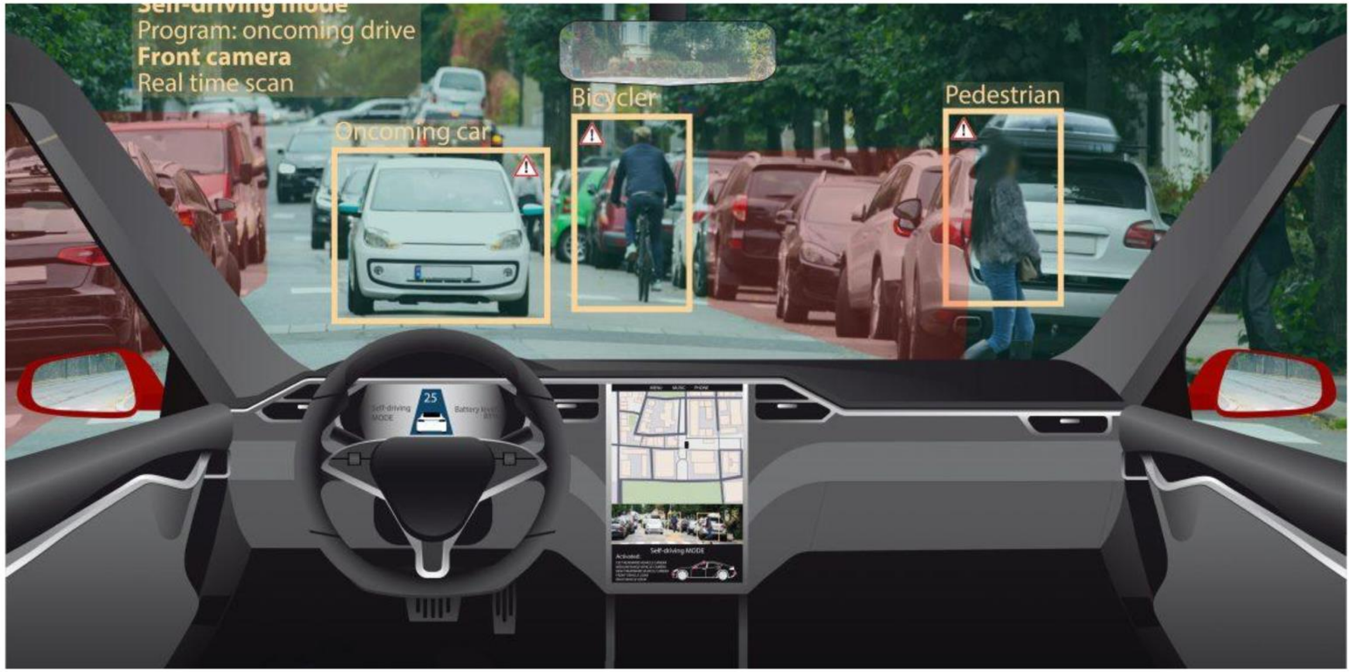
569 rows × 30 columns

P_M Probability of Malignant

P_B Probability of Benign

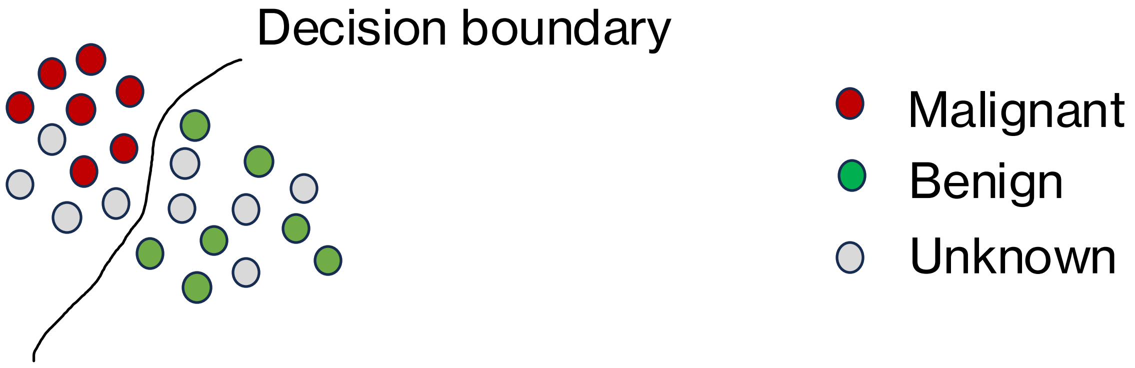
The concept of uncertainty

- Uncertainty is defined the extend to which the machine is confident in its predictions.
- It is important to compute ML uncertainty to support reasoning stages



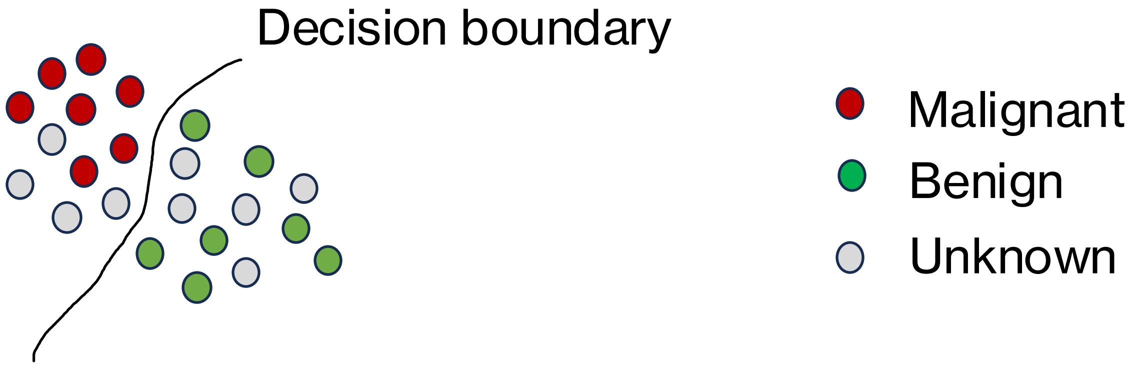
The concept of uncertainty

- All data instances are not equally informative



The concept of uncertainty

- All data instances are not equally informative



- Instances closer to the decision boundary can significantly change the shape of the decision boundary.
- All data instances can not necessarily improve model uncertainty.

Source of uncertainty

- Uncertainty in Machine Learning can come from two sources

Let us consider the cross-entropy loss

$$log(p(y_i|x_i,\beta)) = -(y_i log\big(p(x_i)\big) + (1-y_i)\, log(\, 1-p(x_i)\,))$$

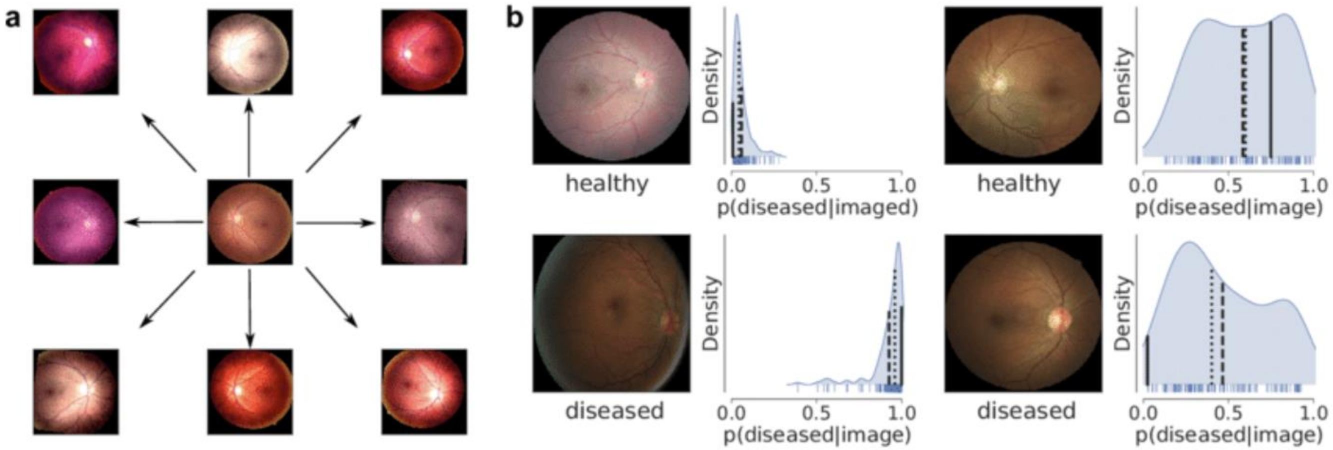
- When model parameters (β) lack sufficient knowledge , it is called **Epistemic** uncertainty
  - Captured by individual predictions

- When observations are noisy and data sufficient knowledge , it is called **Aleatoric** uncertainty
  - Captured by multiple predictions

McMaster University

How do we reduce uncertainty

- **Epistemic** uncertainty can be reduced by:
  - Data Augmentation



Picture Source: Expert-validated estimation of diagnostic uncertainty for deep neural networks in diabetic retinopathy detection Ayhan et. al 2020

How do we reduce uncertainty

- **Epistemic** uncertainty can be reduced by:
  - Data Augmentation
  - Ensemble Methods

- Augment with adversarial training

  $(x_i + \Delta x_i, y_i)$ where $\Delta x_i = -\in sign\ (\nabla_x \log(p(y_i|x_i))$. Essentially introduce minor change in input

- Encourage $p(y_i|x_i;\beta)$ to be similar to $p(y_i|x_i + \Delta x_i;\beta)$

- Train N models as an ensemble with random initialization

- Combine predictive responses by averaging $p(y|x) = \frac{1}{N}\sum_{i=1}^{N} p(y|x,\ \beta_i)$

How do we reduce uncertainty

- **Epistemic** uncertainty can be reduced by:
  - Data Augmentation
  - Ensemble Methods

- **Aleatoric** uncertainty can be reduced by:
  - Reducing our assumptions about the correct model that generated the data.

Active Learning

- **Active learning** is a special case of machine learning in which a learning algorithm can interactively query an oracle (machine or human) to provide annotation on un-labelled or noisy labelled data instance.

- This label is fed back into the machine learning model, and new decision boundary is learned.

- A reasonable strategy is to request feedback from the oracle on instance which might be most informative.

Active Learning

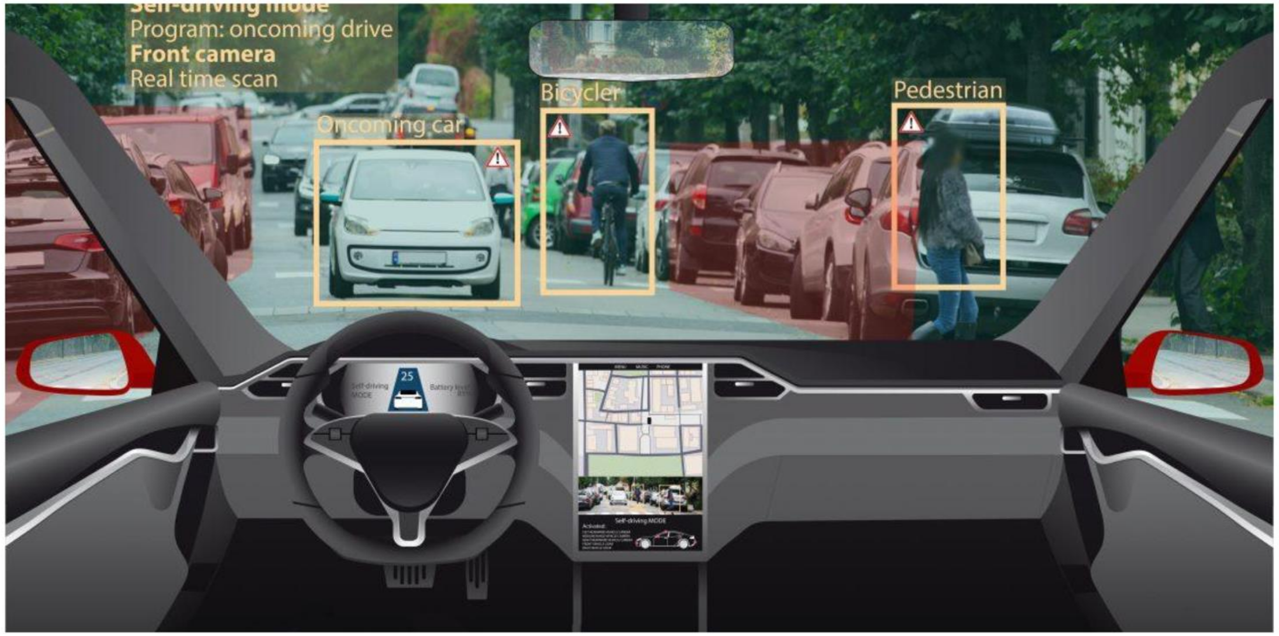
- **Active learning** is a special case of machine learning in which a learning algorithm can interactively query an oracle (machine or human) to provide annotation on un-labelled or noisy labelled data instance.

- An algorithm for uncertainty sampling from a finite training set using a single classifier.

1. Obtain an initial classifier
2. While expert is willing to label instances
   (a) Apply the current classifier to each unlabeled instance
   (b) Find the b instances for which the classifier is least certain of class membership
   (c) Have the expert label the subsample of b instances
   (d) Train a new classifier on all labeled instances

Heterogeneous Uncertainty Sampling for Supervised Learning, David and Jason 1994

How do we measure uncertainty



$P_{cyclist}$ = 0.46
$P_{pedestrian}$ = 0.32
P_{sign} = 0.19
P_{animal} = 0.02

Measuring uncertainty

- Difference between 100% confidence and most confidently predicted labels.

Least Confidence Sampling $\left(1 - P_{cyclist}\right) * \frac{k}{k-1}$

k = number of classes

Uncertainty score for this instance = (1 - 0.46)* 4/3 = 0.72

Uncertainty score for next instance = (1 - 0.68)* 4/3 = 0.42

This instance

$P_{cyclist}$ = 0.46
$P_{pedestrian}$ = 0.32
P_{sign} = 0.19
P_{animal} = 0.02

Next instance

$P_{cyclist}$ = 0.68
$P_{pedestrian}$ = 0.11
P_{sign} = 0.19
P_{animal} = 0.02

McMaster University

Active Learning

- **Active learning** is a special case of machine learning in which a learning algorithm can interactively query an oracle (machine or human) to provide annotation on un-labelled or noisy labelled data instance.

- An algorithm for uncertainty sampling from a finite training set using a single classifier.

1. Obtain an initial classifier
2. While expert is willing to label instances
   (a) Apply the current classifier to each unlabeled instance
   (b) Find the b instances for which the classifier is least certain of class membership
   (c) Have the expert label the subsample of b instances
   (d) Train a new classifier on all labeled instances

We may compute uncertainty in different ways

Algorithm Source: Heterogeneous Uncertainty Sampling for Supervised Learning, David and Jason 1994

McMaster University

Measuring uncertainty

- Difference between two most confident predicted labels.

Margin of Confidence Sampling = $\left(1 - P_{cyclist} - P_{pedestrian}\right)$

This instance

$P_{cyclist}$ = 0.46
$P_{pedestrian}$ = 0.32
P_{sign} = 0.19
P_{animal} = 0.02

Next instance

$P_{cyclist}$ = 0.68
$P_{pedestrian}$ = 0.11
P_{sign} = 0.19
P_{animal} = 0.02

Uncertainty score for this instance = (1 - 0.46 – 0.32)= 0.22

Uncertainty score for next instance = (1 - 0.68 – 0.19) = 0.13

Measuring uncertainty

- Difference between two most confident predicted labels.

Ratio of Confidence Sampling = $\left(\frac{P_{cyclist}}{P_{pedestrian}}\right)$

This instance

$P_{cyclist}$ = 0.46
$P_{pedestrian}$ = 0.32
P_{sign} = 0.19
P_{animal} = 0.02

Next instance

$P_{cyclist}$ = 0.68
$P_{pedestrian}$ = 0.19
P_{sign} = 0.11
P_{animal} = 0.02

Uncertainty score for this instance = (0.46 / 0.32)= 1.44

Uncertainty score for next instance = (0.68 / 0.19) = 3.57

Measuring uncertainty

- The entropy gives a measure of uncertainty about the actual structure of a system.
- The information content of an outcome specifying $\mathrm{X} = x_i$ is defined, $log_a\left(\frac{1}{P_i}\right)$
- The information entropy of a message is defined as the expected average amount of information to be conveyed about X

Source: Entropy-based Sampling Approaches for Multi-Class Imbalanced Problems , Lusi et. al. 2020

Measuring uncertainty

- The entropy gives a measure of uncertainty about the actual structure of a system.
- The information content of an outcome specifying $\mathrm{X} = x_i$ is defined, $log_a\left(\frac{1}{P_i}\right)$
- The information entropy of a message is defined as the expected average amount of information to be conveyed about X

Measuring uncertainty

- The entropy gives a measure of uncertainty about the actual structure of a system.
- The information content of an outcome specifying $\mathrm{X} = x_i$ is defined, $log_a\left(\frac{1}{P_i}\right)$
- The information entropy of a message is defined as the expected average amount of information to be conveyed about X

Entropy based Sampling $= \frac{-\sum_{i=0}^{k} P_i \log(P_i)}{\log(k)}$

Uncertainty score for this instance = -0.80

This instance

$P_{cyclist}$ = 0.46
$P_{pedestrian}$ = 0.32
P_{sign} = 0.19
P_{animal} = 0.02

Measuring uncertainty

- The entropy gives a measure of uncertainty about the actual structure of a system.
- The information content of an outcome specifying $X = x_i$ is defined, $log_a\left(\frac{1}{P_i}\right)$
- The information entropy of a message is defined as the expected average amount of information to be conveyed about X

Entropy based Sampling $= \frac{-\sum_{i=0}^{k} P_i \log(P_i)}{\log(k)}$

This instance

$P_{cyclist}$ = 0.46
$P_{pedestrian}$ = 0.32
P_{sign} = 0.19
P_{animal} = 0.02

Next instance

$P_{cyclist}$ = 0.68
$P_{pedestrian}$ = 0.19
P_{sign} = 0.11
P_{animal} = 0.02

Uncertainty score for this instance = -0.80

Uncertainty score for next instance = -0.64

McMaster University

Measuring uncertainty

- The entropy gives a measure of uncertainty about the actual structure of a system.

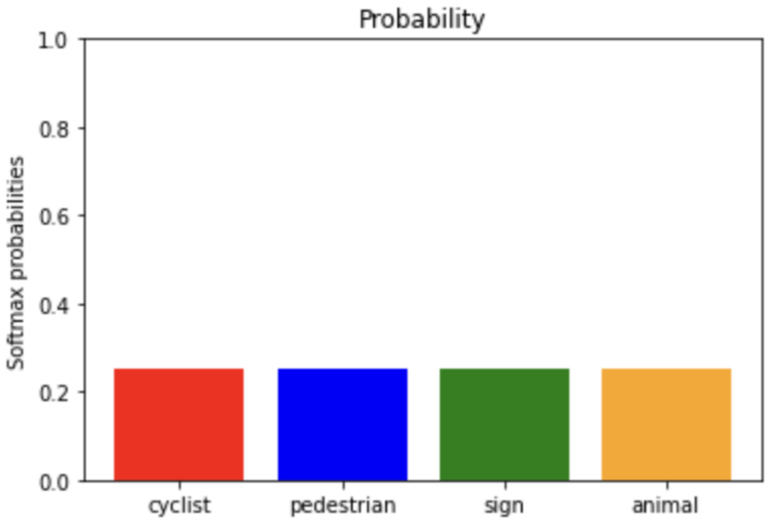
$$\text{Entropy based Sampling} = \frac{-\sum_{i=0}^{k} P_i \log(P_i)}{\log(k)}$$

Measuring uncertainty

- The entropy gives a measure of uncertainty about the actual structure of a system.

Entropy based Sampling $= \frac{-\sum_{i=0}^{k} P_i \log(P_i)}{\log(k)}$

Less variability, all equally likely events



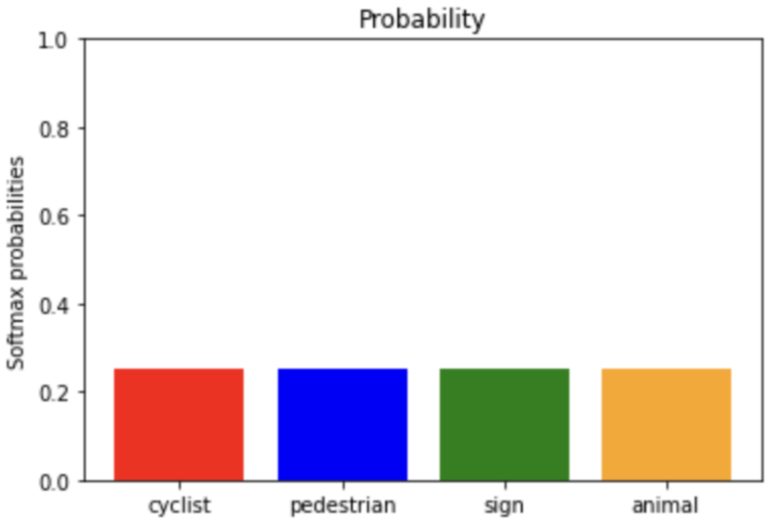
High Entropy

Measuring uncertainty

- The entropy gives a measure of uncertainty about the actual structure of a system.

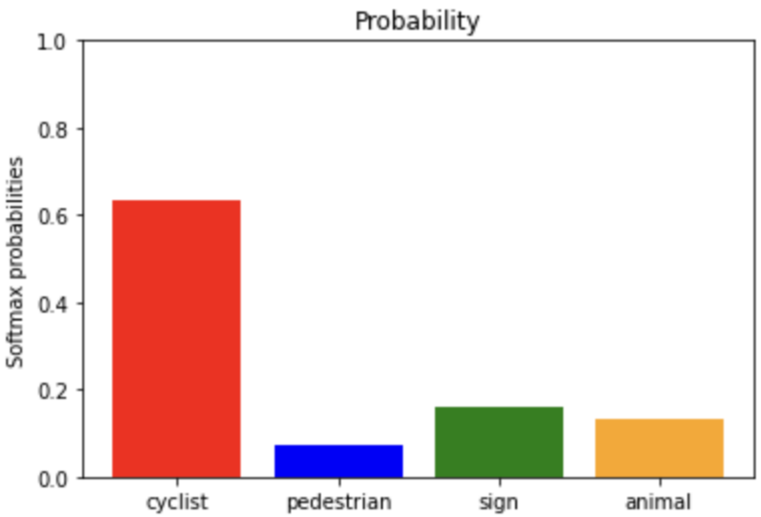
$$\text{Entropy based Sampling} = \frac{-\sum_{i=0}^{k} P_i \log(P_i)}{\log(k)}$$

Less variability, all equally likely events



High Entropy

More variability, 3 highly unlikely events



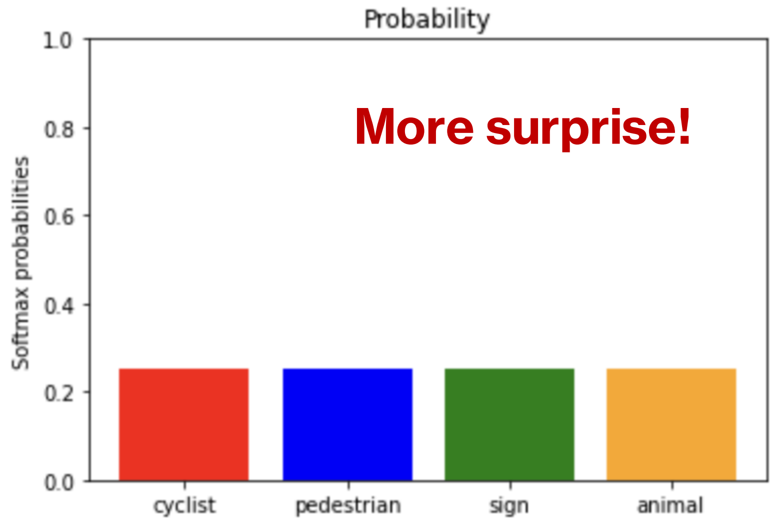
Low Entropy

Measuring uncertainty

- The entropy gives a measure of uncertainty about the actual structure of a system.

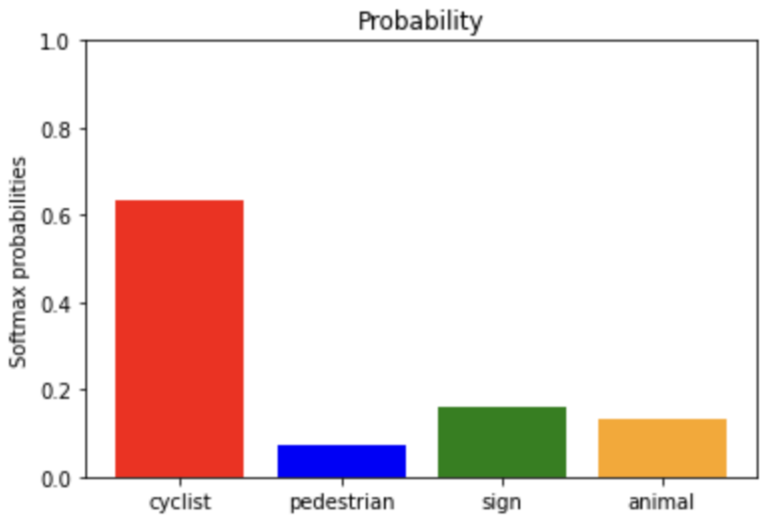
Entropy based Sampling $= \frac{-\sum_{i=0}^{k} P_i \log(P_i)}{\log(k)}$

Less variability, all equally likely events



High Entropy

More variability, 3 highly unlikely events



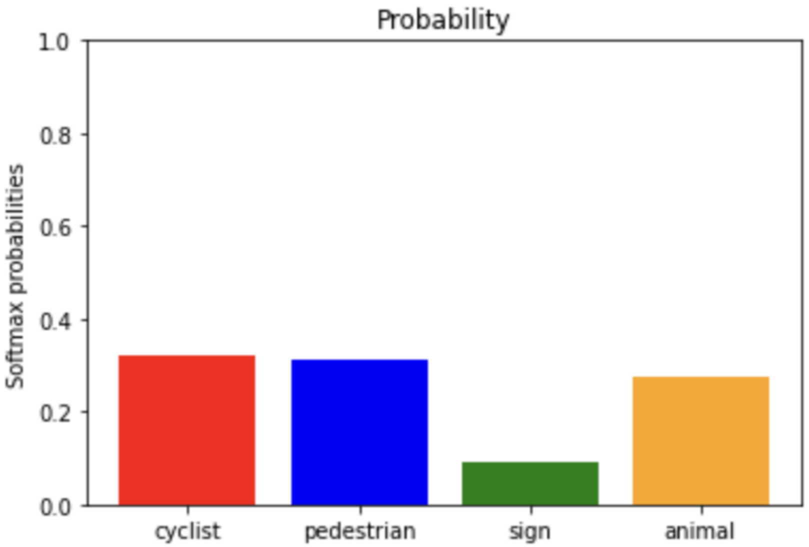
Low Entropy

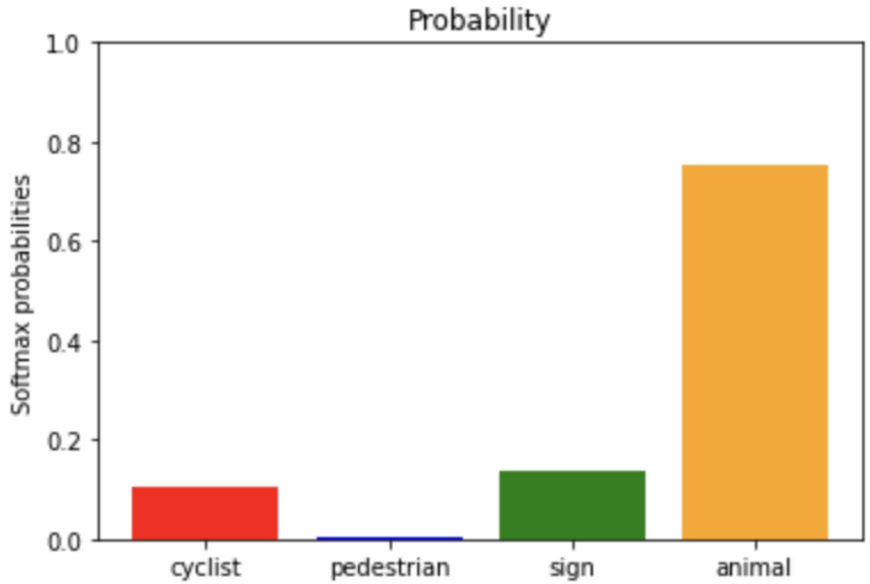
McMaster University

Measuring uncertainty

- Which one is high entropy and low entropy distributions below?

Entropy based Sampling $= \frac{-\sum_{i=0}^{k} P_i \log(P_i)}{\log(k)}$

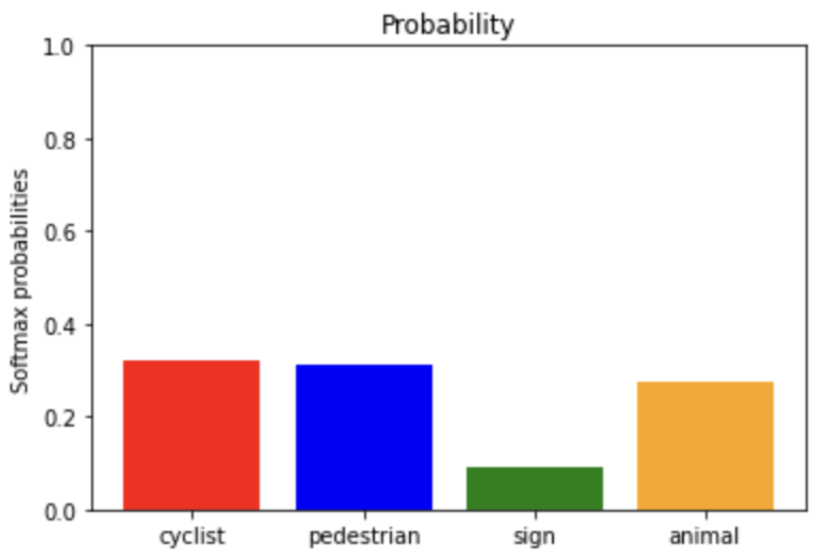




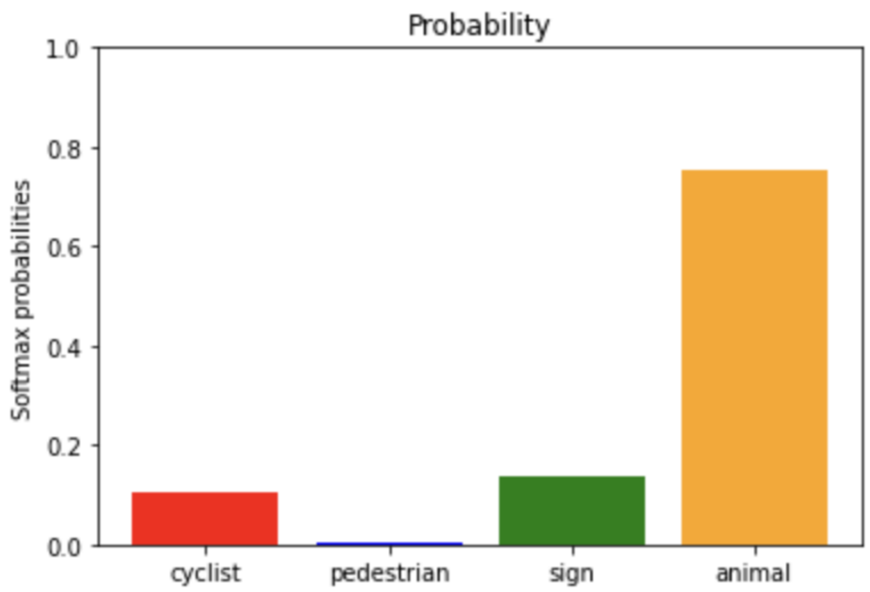
Measuring uncertainty

- Which one is high entropy and low entropy distributions below?

Entropy based Sampling $= \dfrac{-\sum_{i=0}^{k} P_i \log(P_i)}{\log(k)}$



High Entropy



Low Entropy

Measuring uncertainty

- Interpreting entropy?
  - If the probability is 1.0, the model is completely predictable, thus no entropy.
  - If the probability is 0.0, the model does not know anything about the data point. In other words, there is no contribution of entropy to such individuals

Sampling Strategies in Active Learning

1. Obtain an initial classifier
2. While expert is willing to label instances
   (a) Apply the current classifier to each unlabeled instance
   (b) Find the b instances for which the classifier is least certain of class membership
   (c) Have the expert label the subsample of b instances
   (d) Train a new classifier on all labeled instances

This is a sampling step

Sampling Strategies in Active Learning

1. Obtain an initial classifier
2. While expert is willing to label instances
   (a) Apply the current classifier to each unlabeled instance
   (b) Find the b instances for which the classifier is least certain of class membership
   (c) Have the expert label the subsample of b instances
   (d) Train a new classifier on all labeled instances

In which order should this be presented?

Sampling Strategies in Active Learning

- Uncertainty based sampling
  - Least Confidence Sampling captures how confident the most confident prediction is.
  - Margin of Confidence Sampling captures difference between two most confident predictions
  - Ratio of Confidence Sampling captures proportion of two most confident predictions.
  - Entropy based sampling captures how much every confidence differs.

Sampling Strategies in Active Learning

1. Obtain an initial classifier
2. While expert is willing to label instances
   (a) Apply the current classifier to each unlabeled instance
   (b) Find the b instances for which the classifier is least certain of class membership
   (c) Have the expert label the subsample of b instances
   (d) Train a new classifier on all labeled instances

In which order should this be presented

Sampling Strategies in Active Learning

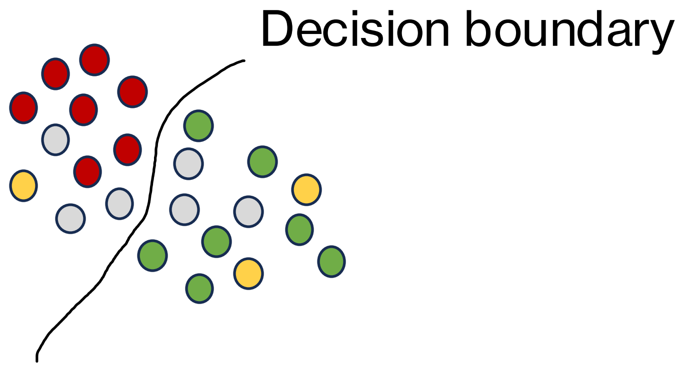
1. Obtain an initial classifier
2. While expert is willing to label instances
   (a) Apply the current classifier to each unlabeled instance
   (b) Find the b instances for which the classifier is least certain of class membership
   (c) Have the expert label the subsample of b instances
   (d) Train a new classifier on all labeled instances

In which order should this be presented

For instance, rank instances based uncertainty scores

Sampling Strategies in Active Learning

- Uncertainty based sampling
  - Least Confidence Sampling captures how confident the most confident prediction is.
  - Margin of Confidence Sampling captures difference between two most confident predictions
  - Ratio of Confidence Sampling captures proportion of two most confident predictions.
  - Entropy based sampling captures how much every confidence differs.
- Diversity Based Sampling



Sampling Strategies in Active Learning

- Uncertainty based sampling
  - Least Confidence Sampling captures how confident the most confident prediction is.
  - Margin of Confidence Sampling captures difference between two most confident predictions
  - Ratio of Confidence Sampling captures proportion of two most confident predictions.
  - Entropy based sampling captures how much every confidence differs.
- Diversity Based Sampling
  - Sampling training samples that are maximally different(outliers) from training set
  - **Model-based outlier sampling** –
    - Determine which instances are unknown to the model in its current state

Sampling Strategies in Active Learning

- Model-based outlier sampling

	mean radius	mean texture	mean perimeter	mean area	mean smoothness	mean compactness	mean concavity	mean concave points	mean symmetry	mean fractal dimension	...	worst radius	worst texture	worst perimeter
0	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.30010	0.14710	0.2419	0.07871	...	25.380	17.33	184.60
1	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.08690	0.07017	0.1812	0.05667	...	24.990	23.41	158.80
2	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.19740	0.12790	0.2069	0.05999	...	23.570	25.53	152.50
3	11.42	20.38	77.58	386.1	0.14250	0.28390	0.24140	0.10520	0.2597	0.09744	...	14.910	26.50	98.87
4	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.19800	0.10430	0.1809	0.05883	...	22.540	16.67	152.20
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
564	21.56	22.39	142.00	1479.0	0.11100	0.11590	0.24390	0.13890	0.1726	0.05623	...	25.450	26.40	166.10
565	20.13	28.25	131.20	1261.0	0.09780	0.10340	0.14400	0.09791	0.1752	0.05533	...	23.690	38.25	155.00
566	16.60	28.08	108.30	858.1	0.08455	0.10230	0.09251	0.05302	0.1590	0.05648	...	18.980	34.12	126.70
567	20.60	29.33	140.10	1265.0	0.11780	0.27700	0.35140	0.15200	0.2397	0.07016	...	25.740	39.42	184.60
568	7.76	24.54	47.92	181.0	0.05263	0.04362	0.00000	0.00000	0.1587	0.05884	...	9.456	30.37	59.16

569 rows × 30 columns

Instances that are lowest ranked in distribution

Sampling Strategies in Active Learning

- Uncertainty based sampling
  - Least Confidence Sampling captures how confident the most confident prediction is.
  - Margin of Confidence Sampling captures difference between two most confident predictions
  - Ratio of Confidence Sampling captures proportion of two most confident predictions.
  - Entropy based sampling captures how much every confidence differs.
- Diversity Based Sampling
  - Sampling training samples that are maximally different(outliers) from training set
  - Model-based outlier sampling –
    - Determine which instances are unknown to the model in its current state
  - **Cluster-based outlier sampling** –
    - Identify a diverse selection of instances using clustering algorithms.

Sampling Strategies in Active Learning

- Cluster-based outlier sampling

	mean radius	mean texture	mean perimeter	mean area	mean smoothness	mean compactness	mean concavity	mean concave points	mean symmetry	mean fractal dimension	...	worst radius	worst texture	worst perimeter
0	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.30010	0.14710	0.2419	0.07871	...	25.380	17.33	184.60
1	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.08690	0.07017	0.1812	0.05667	...	24.990	23.41	158.80
2	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.19740	0.12790	0.2069	0.05999	...	23.570	25.53	152.50
3	11.42	20.38	77.58	386.1	0.14250	0.28390	0.24140	0.10520	0.2597	0.09744	...	14.910	26.50	98.87
4	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.19800	0.10430	0.1809	0.05883	...	22.540	16.67	152.20
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
564	21.56	22.39	142.00	1479.0	0.11100	0.11590	0.24390	0.13890	0.1726	0.05623	...	25.450	26.40	166.10
565	20.13	28.25	131.20	1261.0	0.09780	0.10340	0.14400	0.09791	0.1752	0.05533	...	23.690	38.25	155.00
566	16.60	28.08	108.30	858.1	0.08455	0.10230	0.09251	0.05302	0.1590	0.05648	...	18.980	34.12	126.70
567	20.60	29.33	140.10	1265.0	0.11780	0.27700	0.35140	0.15200	0.2397	0.07016	...	25.740	39.42	184.60
568	7.76	24.54	47.92	181.0	0.05263	0.04362	0.00000	0.00000	0.1587	0.05884	...	9.456	30.37	59.16

569 rows × 30 columns

Observations that are similar maybe clustered together.

Sampling Strategies in Active Learning

- Uncertainty based sampling
  - Least Confidence Sampling captures how confident the most confident prediction is.
  - Margin of Confidence Sampling captures difference between two most confident predictions
  - Ratio of Confidence Sampling captures proportion of two most confident predictions.
  - Entropy based sampling captures how much every confidence differs.
- Diversity Based Sampling
  - Sampling training samples that are maximally different(outliers) from training set
  - Model-based outlier sampling –
    - Determine which instances are unknown to the model in its current state
  - Cluster-based outlier sampling –
    - Identify a diverse selection of instances using clustering algorithms.
  - **Representative sampling -**
    - Identify samples that are like the target domain.

Sampling Strategies in Active Learning

- Representative sampling

	mean radius	mean texture	mean perimeter	mean area	mean smoothness	mean compactness	mean concavity	mean concave points	mean symmetry	mean fractal dimension	...	worst radius	worst texture	worst perimeter
0	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.30010	0.14710	0.2419	0.07871	...	25.380	17.33	184.60
1	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.08690	0.07017	0.1812	0.05667	...	24.990	23.41	158.80
2	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.19740	0.12790	0.2069	0.05999	...	23.570	25.53	152.50
3	11.42	20.38	77.58	386.1	0.14250	0.28390	0.24140	0.10520	0.2597	0.09744	...	14.910	26.50	98.87
4	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.19800	0.10430	0.1809	0.05883	...	22.540	16.67	152.20
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
564	21.56	22.39	142.00	1479.0	0.11100	0.11590	0.24390	0.13890	0.1726	0.05623	...	25.450	26.40	166.10
565	20.13	28.25	131.20	1261.0	0.09780	0.10340	0.14400	0.09791	0.1752	0.05533	...	23.690	38.25	155.00
566	16.60	28.08	108.30	858.1	0.08455	0.10230	0.09251	0.05302	0.1590	0.05648	...	18.980	34.12	126.70
567	20.60	29.33	140.10	1265.0	0.11780	0.27700	0.35140	0.15200	0.2397	0.07016	...	25.740	39.42	184.60
568	7.76	24.54	47.92	181.0	0.05263	0.04362	0.00000	0.00000	0.1587	0.05884	...	9.456	30.37	59.16

569 rows × 30 columns

Instances that might represent the target domain

Sampling Strategies in Active Learning

- Uncertainty based sampling
  - Least Confidence Sampling captures how confident the most confident prediction is.
  - Margin of Confidence Sampling captures difference between two most confident predictions
  - Ratio of Confidence Sampling captures proportion of two most confident predictions.
  - Entropy based sampling captures how much every confidence differs.
- Diversity Based Sampling
  - Sampling training samples that are maximally different(outliers) from training set
  - Model-based outlier sampling –
    - Determine which instances are unknown to the model in its current state
  - Cluster-based outlier sampling –
    - Identify a diverse selection of instances using clustering algorithms.
  - Representative sampling -
    - identify a sample that looks like the target domain.
  - **Real-world diversity** –
    - Identify samples that may have been missed for certain demographics.

Sampling Strategies in Active Learning

- Real-world diversity sampling

	mean radius	mean texture	mean perimeter	mean area	mean smoothness	mean compactness	mean concavity	mean concave points	mean symmetry	mean fractal dimension	...	worst radius	worst texture	worst perimeter
0	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.30010	0.14710	0.2419	0.07871	...	25.380	17.33	184.60
1	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.08690	0.07017	0.1812	0.05667	...	24.990	23.41	158.80
2	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.19740	0.12790	0.2069	0.05999	...	23.570	25.53	152.50
3	11.42	20.38	77.58	386.1	0.14250	0.28390	0.24140	0.10520	0.2597	0.09744	...	14.910	26.50	98.87
4	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.19800	0.10430	0.1809	0.05883	...	22.540	16.67	152.20
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
564	21.56	22.39	142.00	1479.0	0.11100	0.11590	0.24390	0.13890	0.1726	0.05623	...	25.450	26.40	166.10
565	20.13	28.25	131.20	1261.0	0.09780	0.10340	0.14400	0.09791	0.1752	0.05533	...	23.690	38.25	155.00
566	16.60	28.08	108.30	858.1	0.08455	0.10230	0.09251	0.05302	0.1590	0.05648	...	18.980	34.12	126.70
567	20.60	29.33	140.10	1265.0	0.11780	0.27700	0.35140	0.15200	0.2397	0.07016	...	25.740	39.42	184.60
568	7.76	24.54	47.92	181.0	0.05263	0.04362	0.00000	0.00000	0.1587	0.05884	...	9.456	30.37	59.16

569 rows × 30 columns

Capture diversity of demographics

Gender	Age
F	30
M	40
F	50
M	60
F	70

Sampling Strategies in Active Learning

- Uncertainty based sampling
  - Least Confidence Sampling captures how confident the most confident prediction is.
  - Margin of Confidence Sampling captures difference between two most confident predictions
  - Ratio of Confidence Sampling captures proportion of two most confident predictions.
  - Entropy based sampling captures how much every confidence differs.
- Diversity Based Sampling
  - Sampling training samples that are maximally different(outliers) from training set
  - Model-based outlier sampling –
    - Determine which instances are unknown to the model in its current state
  - Cluster-based outlier sampling –
    - Identify a diverse selection of instances using clustering algorithms.
  - Representative sampling -
    - identify a sample that looks like the target domain.
  - Real-world diversity –
    - Identify samples that may have been missed for certain demographics.

Sampling Strategies in Active Learning

- Uncertainty based sampling
  - Least Confidence Sampling captures how confident the most confident prediction is.
  - Margin of Confidence Sampling captures difference between two most confident predictions
  - Ratio of Confidence Sampling captures proportion of two most confident predictions.
  - Entropy based sampling captures how much every confidence differs.
- Diversity Based Sampling
  - Sampling training samples that are maximally different(outliers) from training set
  - Model-based outlier sampling –
    - Determine which instances are unknown to the model in its current state
  - Cluster-based outlier sample –
    - Identify a diverse selection of instances using clustering algorithms.
  - Representative sampling -
    - identify a sample that looks like the target domain.
  - Real-world diversity –
    - Identify samples that may have been missed for certain demographics.

How do we measure outliers?

Future Lecture

- We will cover this after we learn a few clustering and dimensionality reduction algorithms.

How do we measure outliers?

Readings

Required Readings:

- None

***Supplemental Readings*:**

- Human-in-the-loop Machine Learning , Robert Monarch
- Linked paper sources on slides

Thank You

McMaster University