CS2109S Tutorial 4

Classification and Logistic Regression

(AY 25/26 Semester 2)

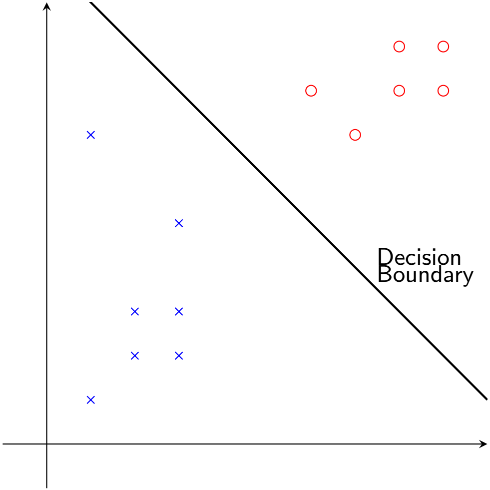
February 19, 2026

(Prepared by Benson)

Contents

Recap: Classification

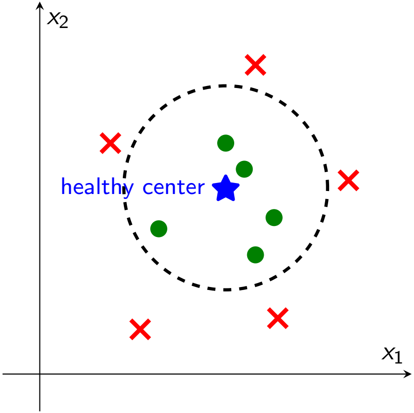
Target: Solve **classification** tasks.



Q1. Tumor Classification

(a) Describe the geometric pattern exhibited by the two classes.

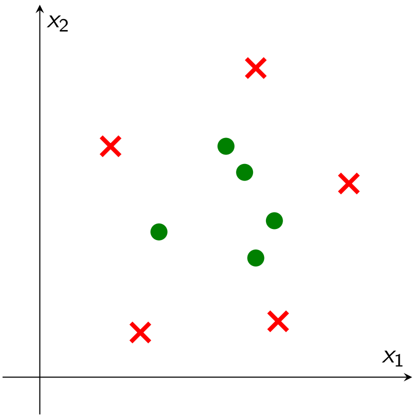
- Concentric pattern where distance from the center correlates with tumor type.



Q1. Tumor Classification

(b) What if we train a logistic regression classifier using the feature space $[x_1, x_2]$?

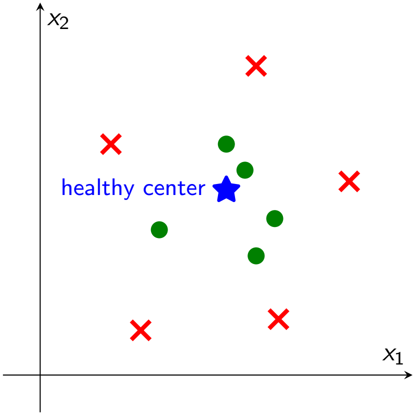
- The two classes are not **linearly separable**.
- The model will fail to classify all 10 samples correctly.

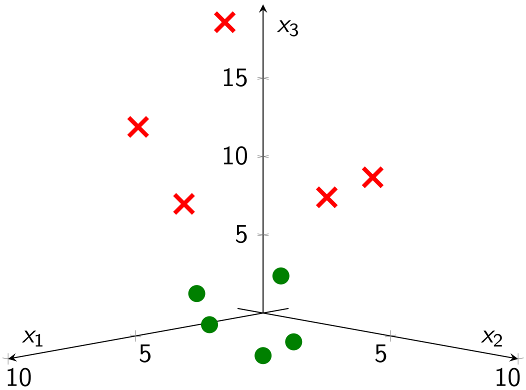


Q1. Tumor Classification

(c) Add a new transformed feature $x_3 = (x_1 - 5)^2 + (x_2 - 5)^2$. What does this mean?

- The distance to the healthy center. Exactly what we need!





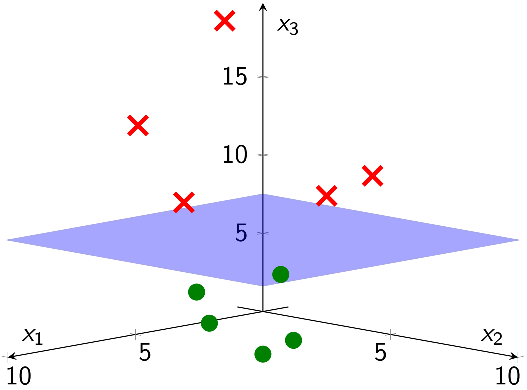
Q1. Tumor Classification

(d) Suggest a decision boundary in the new feature space $[x_1, x_2, x_3]$.

- $x_3 = 7.5$ is a possible decision boundary that perfectly classifies the data.

Then, classify a new sample $(x_1, x_2) = (6.0, 7.5)$.

- $x_3 = (6.0 - 5.0)^2 + (7.5 - 5.0)^2 = 7.25 < 7.5$, so we classify as Benign.

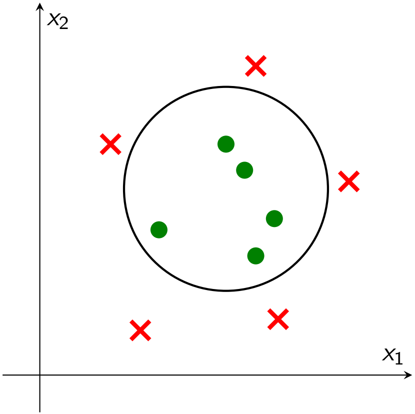


Sample	x_1	x_2	x_3	Label
A	5.5	5.5	0.50	Benign
B	5.0	6.2	1.44	Benign
C	3.2	3.9	4.45	Benign
D	5.8	3.2	3.88	Benign
E	6.3	4.2	2.33	Benign
F	8.3	5.2	10.93	Malignant
G	5.8	8.3	11.53	Malignant
H	1.9	6.2	11.05	Malignant
I	2.7	1.2	19.73	Malignant
J	6.4	1.5	14.21	Malignant

Q1. Tumor Classification

(e) Draw the decision boundary back in the original plot.

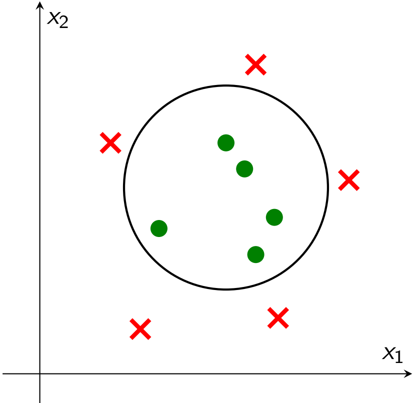
- $(x_1 - 5)^2 + (x_2 - 5)^2 = 7.5$.
- This is a circle! (i.e. non-linear decision boundary)



Discussion Question

The transformed feature x_3 added in this question is sort of idealized – we already "guessed" the pattern of the data correctly. In a practical scenario, if we realise this dataset is not linearly separable, what transformed features will we add instead?

- Usually we will add x_1^2 and x_2^2.
- The model will learn the correct weights (coefficients) to produce the same circle.



Recap: Logistic Regression

Target: Solve **classification** tasks (output a **probability** between 0 and 1).

- First Try (regression): $p = \boldsymbol{w} \cdot \boldsymbol{x}$
  - FAIL: $\boldsymbol{w} \cdot \boldsymbol{x}$ can be outside $[0, 1]$.
- Second Try (odds ratio): $\dfrac{p}{1-p} = \boldsymbol{w} \cdot \boldsymbol{x}$
  - FAIL: $\dfrac{p}{1-p}$ is always non-negative, but $\boldsymbol{w} \cdot \boldsymbol{x}$ can be outside $[0, \infty)$.
- Third Try (logits): $\log\left(\dfrac{p}{1-p}\right) = \boldsymbol{w} \cdot \boldsymbol{x}$ (works!)
  - Rearrange terms: $p = \dfrac{1}{1 + e^{-\boldsymbol{w} \cdot \boldsymbol{x}}}$

Recap: Logistic Regression

Hypothesis:

$$h_{\boldsymbol{w}}(x) = \frac{1}{1 + e^{-\boldsymbol{w}^\top \boldsymbol{x}}} = \sigma(\boldsymbol{w}^\top \boldsymbol{x})$$

$$\boxed{\sigma(z) = \frac{1}{1 + e^{-z}}}$$

Loss function:

- **Mean Squared Error** is not convex under logistic regression $\Rightarrow$ gradient descent might not reach global minimum.
- Use **binary cross entropy loss** instead:

"surprisal" in entropy!

$$BCE(p) = \begin{cases} -\log(p) & \text{if } y = 1 \\ -\log(1-p) & \text{if } y = 0 \end{cases}$$
$$= -y\log(p) - (1-y)\log(1-p)$$

Q2. Loss function of Logistic Regression

Hypothesis:

$$h_{\boldsymbol{w}}(x) = p(y = 1 \mid x) = \sigma(\boldsymbol{w}^\top \boldsymbol{x})$$

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

(a) Show that $p = \sigma(z)$ where $z = \ln\left(\frac{p}{1-p}\right)$. log odds ratio!

$$\begin{aligned}
z &= \ln\left(\frac{p}{1-p}\right) \\
e^z &= \frac{p}{1-p} \\
(1-p)e^z &= p \\
e^z &= p(e^z + 1) \\
p &= \frac{e^z}{e^z + 1} = \frac{1}{1 + e^{-z}}
\end{aligned}$$

Q2. Loss Function of Logistic Regression

Derivative of Sigmoid Function

Let $\sigma(z) = \dfrac{1}{1+e^{-z}}$. We have $\sigma'(z) = \sigma(z)\,(1-\sigma(z))$.

(b) Show that $\frac{\partial}{\partial w_j} BCE(y, \sigma(z)) = (\sigma(z) - y)x_j$, where

$$BCE(y, \sigma(z)) = -y \log \sigma(z) - (1-y)\log(1-\sigma(z))$$

Chain Rule; $p = \sigma(\boldsymbol{w}^\top \boldsymbol{x})$; $\boldsymbol{w}^\top \boldsymbol{x} = \boldsymbol{w} \cdot \boldsymbol{x} = \sum w_j x_j$

$$\frac{\partial \log(p)}{\partial w_j} = \underbrace{\frac{1}{p}}_{\frac{\partial \log(p)}{\partial p}} \cdot \frac{\partial p}{\partial w_j} = \frac{1}{p} \cdot \underbrace{p(1-p)}_{\frac{\partial p}{\partial (\boldsymbol{w}^\top \boldsymbol{x})}} \cdot \underbrace{\frac{\partial}{\partial w_j}(\boldsymbol{w}^\top \boldsymbol{x})}_{= x_j} = \boxed{(1-p)x_j}$$

Chain Rule

$$\frac{\partial \log(1-p)}{\partial w_j} = \underbrace{\frac{-1}{1-p}}_{\frac{\partial \log(1-p)}{\partial p}} \cdot \frac{\partial p}{\partial w_j} = \frac{-1}{1-p} \cdot p(1-p) \cdot \underbrace{\frac{\partial}{\partial w_j}(\boldsymbol{w}^\top \boldsymbol{x})}_{= x_j} = \boxed{-p x_j}$$

Q2. Loss Function of Logistic Regression

(b) Show that $\frac{\partial}{\partial w_j} BCE(y, \sigma(z)) = (\sigma(z) - y)x_j$, where

$$BCE(y, \sigma(z)) = -y \log \sigma(z) - (1 - y) \log(1 - \sigma(z))$$

$$\begin{aligned}
\frac{\partial}{\partial w_j} BCE(y, p) &= -y \frac{\partial \log(p)}{\partial w_j} - (1 - y) \frac{\partial \log(1 - p)}{\partial w_j} \\
&= -y(1 - p)x_j - (1 - y)(-px_j) \\
&= -yx_j + ypx_j + px_j - ypx_j \\
&= (p - y)x_j
\end{aligned}$$

Q2. Loss Function of Logistic Regression

Logistic Regression:

$$z = \boldsymbol{w}^\top \boldsymbol{x} \quad \Rightarrow \quad p = \frac{1}{1 + e^{-z}}$$

(Intuition: z predicts $\ln\left(\frac{p}{1-p}\right)$.)

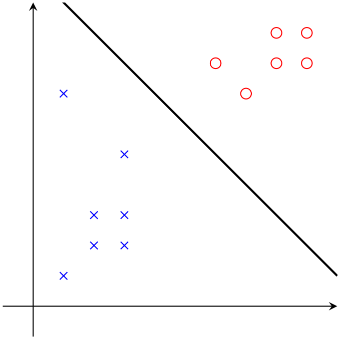
Softmax function (for multi-class classification with k classes):

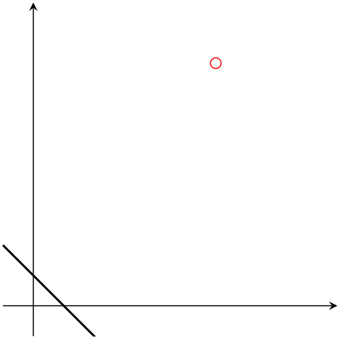
$$z_1, z_2, \ldots, z_k \quad \Rightarrow \quad p_i = \frac{e^{z_i}}{e^{z_1} + e^{z_2} + \cdots + e^{z_k}}$$

(Intuition: z_i predicts $\ln\left(\frac{p_i}{p_k}\right)$.)

Discussion Question

Are probabilities outputted by logistic regression models trustable? In other words, does the p outputted by the model resemble $p(y = 1 \mid \mathbf{x}, \text{training data})$ correctly (or match our intuitions of $p(y = 1)$)?

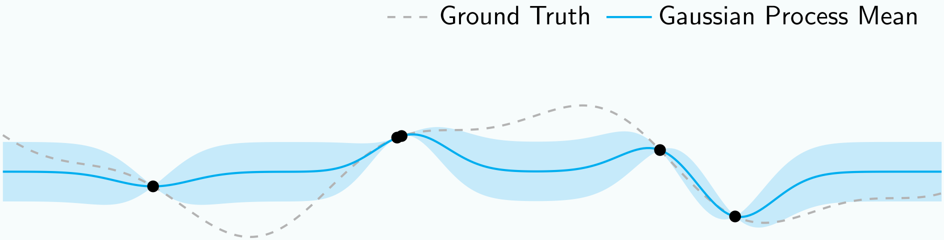




Discussion Question

Bayesian approaches: Models uncertainty by injecting a **prior**.

$$\underbrace{p(\text{belief} \mid \text{data})}_{\text{posterior}} \propto \underbrace{p(\text{data} \mid \text{belief})}_{\text{likelihood}} \times \underbrace{p(\text{belief})}_{\text{prior}}$$



Example: Bayesian (Gaussian Process) Regression

Discussion Question

Demo (if we have time) – See the HTML version.

Q3. Logistic Regression for Multi-Class Classification

(a) Compute the probability of an animal belonging to a certain class and classify them accordingly.

$$\mathbf{w}_{cat} = \begin{bmatrix} 4.2 & -0.01 & -0.12 \end{bmatrix}^\top$$
$$\mathbf{w}_{horse} = \begin{bmatrix} -20 & -0.08 & 35 \end{bmatrix}^\top$$
$$\mathbf{w}_{elephant} = \begin{bmatrix} -1250 & 0.82 & 0.9 \end{bmatrix}^\top$$

Weight (kg)	Length (m)
4.2	0.4
720	2.4
2350	5.5

First animal: $\mathbf{x} = \begin{bmatrix} 1 & 4.2 & 0.4 \end{bmatrix}^\top$

- $\mathbf{w}_{cat} \cdot \mathbf{x} = 1 \cdot 4.2 + 4.2 \cdot (-0.01) + 0.4 \cdot (-0.12) = 4.11$
  $$p_{cat} = \frac{1}{1 + e^{-4.11}} = 0.984$$
- $\mathbf{w}_{horse} \cdot \mathbf{x} = -6.336$
  $$p_{horse} = \frac{1}{1 + e^{6.336}} = 0.00177$$
- $\mathbf{w}_{elephant} \cdot \mathbf{x} = -1246.196$
  $$p_{elephant} = \frac{1}{1 + e^{1246.196}} \approx 0$$

Q3. Logistic Regression for Multi-Class Classification

(a) Compute the probability of an animal belonging to a certain class and classify them accordingly.

$$\mathbf{w}_{cat} = \begin{bmatrix} 4.2 & -0.01 & -0.12 \end{bmatrix}^\top$$
$$\mathbf{w}_{horse} = \begin{bmatrix} -20 & -0.08 & 35 \end{bmatrix}^\top$$
$$\mathbf{w}_{elephant} = \begin{bmatrix} -1250 & 0.82 & 0.9 \end{bmatrix}^\top$$

Weight (kg)	Length (m)
4.2	0.4
720	2.4
2350	5.5

Second animal: $\mathbf{x} = \begin{bmatrix} 1 & 720 & 2.4 \end{bmatrix}^\top$

- $\mathbf{w}_{cat} \cdot \mathbf{x} = -3.288$

  $p_{cat} = \dfrac{1}{1 + e^{3.288}} = 0.0360$
- $\mathbf{w}_{horse} \cdot \mathbf{x} = 6.4$

  $p_{horse} = \dfrac{1}{1 + e^{-6.4}} = 0.998$
- $\mathbf{w}_{elephant} \cdot \mathbf{x} = -657.44$

  $p_{elephant} = \dfrac{1}{1 + e^{657.44}} \approx 0$

Q3. Logistic Regression for Multi-Class Classification

(a) Compute the probability of an animal belonging to a certain class and classify them accordingly.

$$\boldsymbol{w}_{cat} = \begin{bmatrix} 4.2 & -0.01 & -0.12 \end{bmatrix}^\top$$
$$\boldsymbol{w}_{horse} = \begin{bmatrix} -20 & -0.08 & 35 \end{bmatrix}^\top$$
$$\boldsymbol{w}_{elephant} = \begin{bmatrix} -1250 & 0.82 & 0.9 \end{bmatrix}^\top$$

Weight (kg)	Length (m)
4.2	0.4
720	2.4
2350	5.5

Third animal: $\boldsymbol{x} = \begin{bmatrix} 1 & 2350 & 5.5 \end{bmatrix}^\top$

- $\boldsymbol{w}_{cat} \cdot \boldsymbol{x} = -19.96$
  $p_{cat} = \dfrac{1}{1 + e^{19.96}} \approx 0$
- $\boldsymbol{w}_{horse} \cdot \boldsymbol{x} = -15.5$
  $p_{horse} = \dfrac{1}{1 + e^{15.5}} \approx 0$
- $\boldsymbol{w}_{elephant} \cdot \boldsymbol{x} = 681.95$
  $p_{elephant} = \dfrac{1}{1 + e^{-681.95}} \approx 1$

Q3. Logistic Regression for Multi-Class Classification

(b) What if we want to extend the classification task to classify other animals? Can we train a new model while keeping the weights of the previous models?

- For an animal that are very distinct with the three animals, we can create a new logistic regression model without changing the previous weights.
- For classifying a new animal that is similar with one of the classes (e.g, classifying a dog), we need to retrain the old models.

Q3. Logistic Regression for Multi-Class Classification

(c) Derive the pairwise decision boundaries (lines) between each pair of classes in the (weight, length) plane using the given $\boldsymbol{w}$'s.

- cat vs horse:

$$\boldsymbol{w}_{cat} \cdot \boldsymbol{x} = \boldsymbol{w}_{horse} \cdot \boldsymbol{x}$$
$$24.2 + 0.07W - 35.12L = 0$$

- cat vs elephant:

$$\boldsymbol{w}_{cat} \cdot \boldsymbol{x} = \boldsymbol{w}_{elephant} \cdot \boldsymbol{x}$$
$$1254.2 - 0.83W - 1.02L = 0$$

- horse vs elephant:

$$\boldsymbol{w}_{horse} \cdot \boldsymbol{x} = \boldsymbol{w}_{elephant} \cdot \boldsymbol{x}$$
$$1230 - 0.90W + 34.1L = 0$$

$$\boldsymbol{w}_{cat} = \begin{bmatrix} 4.2 & -0.01 & -0.12 \end{bmatrix}^\top$$
$$\boldsymbol{w}_{horse} = \begin{bmatrix} -20 & -0.08 & 35 \end{bmatrix}^\top$$
$$\boldsymbol{w}_{elephant} = \begin{bmatrix} -1250 & 0.82 & 0.9 \end{bmatrix}^\top$$

Recap: Performance measures

Evaluation Metric:

- Judges the performance, doesn't care about the process.

[MRQ] Which of the following link(s) are pruned? Shade <u>all</u> that is/are true.

◯ a	◯ b	◯ c	◯ d	◯ e
◯ f	◯ g	◯ h	◯ i	◯ j
◯ k	◯ l	● m ✓	◯ n	◯ o
◯ p	◯ q	● r ✗	◯ s	● t ✓
◯ u	◯ v	◯ w	◯ x	◯ y
◯ z				

0 / 4

Loss Function:

- Helps with model training. Minimized by the optimizer.

[MRQ] Which of the following link(s) are pruned? Shade <u>all</u> that is/are true.

◯ a	◯ b	◯ c	◯ d	◯ e
◯ f	◯ g	◯ h	◯ i	◯ j
◯ k	◯ l	● m ✓	◯ n	◯ o
◯ p	◯ q	● r ✗	◯ s	● t ✓
◯ u	◯ v	◯ w	◯ x	◯ y
◯ z				

25/26 items right 👍 Try again!

Recap: Performance measures

Which of the following are more likely used as evaluation metrics (rather than loss functions)?

A. Binary cross-entropy loss
B. Accuracy
C. Mean-squared error
D. F1 Score