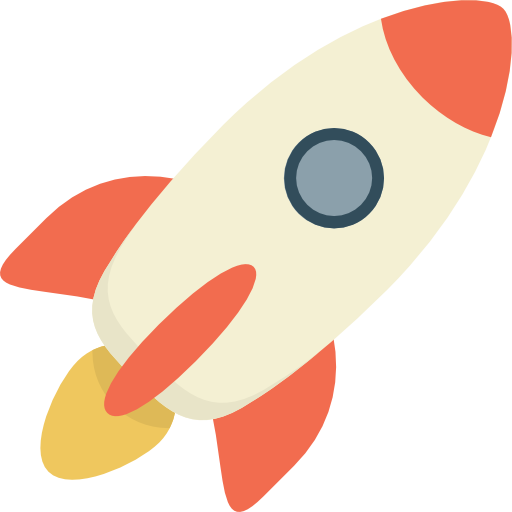
****

**Machine Learning for Product Managers**

Exercise Workbook

|  |  |
| --- | --- |
| **1.** [**Intro to ML**](#kmc7fxgokr40) **2.** [**When to ML**](#uoqn9n4wdytr)  **3.** [**How to ML**](#r5m5hqojzxq) | **4.** [**Get Your Data**](#iml73dvdelw)  **5.** [**Prepare Your Data**](#fe7kr9z1a6kf)  **6.** [**Deploy Your Model**](#hywrv5xb5hao) |

**Section 1: Intro to ML**

**Exercise #1: Choose Your Product (15 minutes)**

|  |  |  |
| --- | --- | --- |
| Product Name |  | |
| What does it do? |  | |
| Where is it available? |  |  |
| Existing ML use cases |  | |
| Your Role |  | |

**Exercise #2a: What type of ML is this? (5 minutes)***Identify the ML problem type for each example provided*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Example** | **Ranking** | **Recommendation** | **Classification** | **Regression** | **Clustering** | **Anomaly Detection** |
| **Google Home**: “Find some good restaurants near me” |  |  |  |  |  |  |
| **Experian:** Gives you a credit score of 780 |  |  |  |  |  |  |
| **Citibank:** Sends you an email about a “suspicious” transaction |  |  |  |  |  |  |
| **Netflix**: Says “Because you watched XXX you may like YYY also” |  |  |  |  |  |  |
| **Google Photos:** Auto-tags a photo as a “beach photo” |  |  |  |  |  |  |
| **Amazon**: Shows “products related” to an item you are viewing |  |  |  |  |  |  |

**Exercise #2b: What type of ML is this? (15 minutes)***Find examples of each ML problem type in the products you use*

|  |  |
| --- | --- |
| **Type** | **Example** |
| Ranking |  |
| Recommendation |  |
| Classification |  |
| Regression |  |
| Clustering |  |
| Anomaly Detection |  |

**Exercise #3: Apply your ML lingo (10 minutes)**

*Your ML model predicts whether a Facebook account is fake or real. Based on the table below, classify aspects of the data as “feature” or “label”*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Facebook Message | # of Likes | Account Gender | # of Friends | Fake or Real |
| One day sale on sunglasses [🕶️](https://emojipedia.org/sunglasses/) | 122 | Male | 2 | Fake |
| Beautiful day for swimming [🏊](https://emojipedia.org/emoji/%F0%9F%8F%8A/) | 64 | Female | 48 | Real |

|  |  |
| --- | --- |
| **Classify this** | **Feature or Label?** |
| Facebook Message |  |
| # of Likes |  |
| Fake |  |
| # of Friends |  |

*Classify each decision made below as a “feature engineering” or “feature selection" activity*

|  |  |
| --- | --- |
| **Classify this** | **Feature Engineering or Feature Selection?** |
| Account gender = Male, Female, Not Known, Not Applicable |  |
| Account gender is not useful for predicting if an account is fake |  |
| Convert message timestamp to UTC format |  |

**Section 2: When to ML**

**Exercise #4: Need Explainability? (5 minutes)**

*Which of the potential ML use cases below need to be explainable?*

|  |  |  |
| --- | --- | --- |
| Use Case | Need Explainability?  (Y/N) | Why?   (Low/Med/High) |
| Amazon recommends a book based on a recent purchase |  | Legal =  Financial = Medical = Ethical = |
| Tesla Auto-pilot decides not to swerve to avoid hitting a small animal for driver safety reasons |  | Legal =  Financial=  Medical =  Ethical = |
| Bank rejects an application for a home mortgage |  | Legal = Financial =  Medical = Ethical = |
| Facebook suggests that your friends tag you in a photo |  | Legal = Financial = Medical =  Ethical = |

**Exercise #5: To ML or not to ML? (15 minutes)**

*Which of these use cases are good candidates for ML? (assuming you have the right skill & resources available)*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Use Case** | **Complex Problem?**  **(Y/N)** | **Changes Over Time?   (Y/N)** | **Need to Scale?  (Y/N)** | **Need 100% Accuracy**  **(Y/N)** | **Needs to be Explainable?**  **(Y/N)** | **Is this Ethical?  (Y/N)** | **Solve Using ML?  (Y/N)** |
| 1) | Which distributors have the greatest sales potential |  |  |  |  |  |  |  |
| 2) | Which products should be sold exclusively to Hispanics in the US |  |  |  |  |  |  |  |
| 3) | Which employees are likely to leave in next 6 months |  |  |  |  |  |  |  |
| 4) | Which resumes should be prioritized for interviewing? |  |  |  |  |  |  |  |
| 5) | Which products need to be protected by copyright laws |  |  |  |  |  |  |  |

**Section 3: How to ML**

**Exercise #6: Test Your Hypothesis (30 minutes)**

1. *Identify a potential ML use case for your chosen product.*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Use Case** | **Complex Problem?**  **(Y/N)** | **Changes Over Time?   (Y/N)** | **Need to Scale?  (Y/N)** | **Need 100% Accuracy**  **(Y/N)** | **Needs to be Explainable?**  **(Y/N)** | **Is this Ethical?  (Y/N)** | **Solve Using ML?  (Y/N)** |
| Predict time needed to prepare food |  |  |  |  |  |  |  |

1. *State your hypothesis using the following format:*

|  |  |
| --- | --- |
| ***Format*** | ***Write Your Hypothesis Here*** |
| *We believe our [target market]*  *Have a problem of [assumption]*  *If we [proposed solution]*  *We can [expected impact]* |  |

1. *Identify risky assumptions:*

|  |  |
| --- | --- |
| Assumption #1 |  |
| Why? |  |
| Why? |  |
| Why? |  |
| Why? |  |
| Why? |  |
| Assumption #2 |  |
| Why? |  |
| Why? |  |
| Why? |  |

1. *Validate risky assumptions:*

|  |  |  |
| --- | --- | --- |
| Assumption #1 | | How to test |
| Why? |  |  |
| Why? |  |
| Why? |  |
| Assumption #2 | | How to test |
| Why? |  |  |
| Why? |  |
| Why? |  |

**Exercise #7: Frame Your ML Problem (20 minutes)**

*a) What do you want your machine learning model to do? (****Tip****: A qualitative statement focused on a real goal.)  
  
The ML model should..*

|  |
| --- |
|  |

*b) Write down the success and failure metrics for the ML system.*

*Our success metrics are..*

|  |
| --- |
|  |

*Our model is deemed a failure if..*

|  |
| --- |
|  |

*c) Write the output that you want your ML model to produce.*

*The output from our ML model will be..*

|  |
| --- |
|  |

*d) Write when the output from the ML model needs to be available & how it will be used.*

*The output will need to be made available when..*

|  |
| --- |
|  |

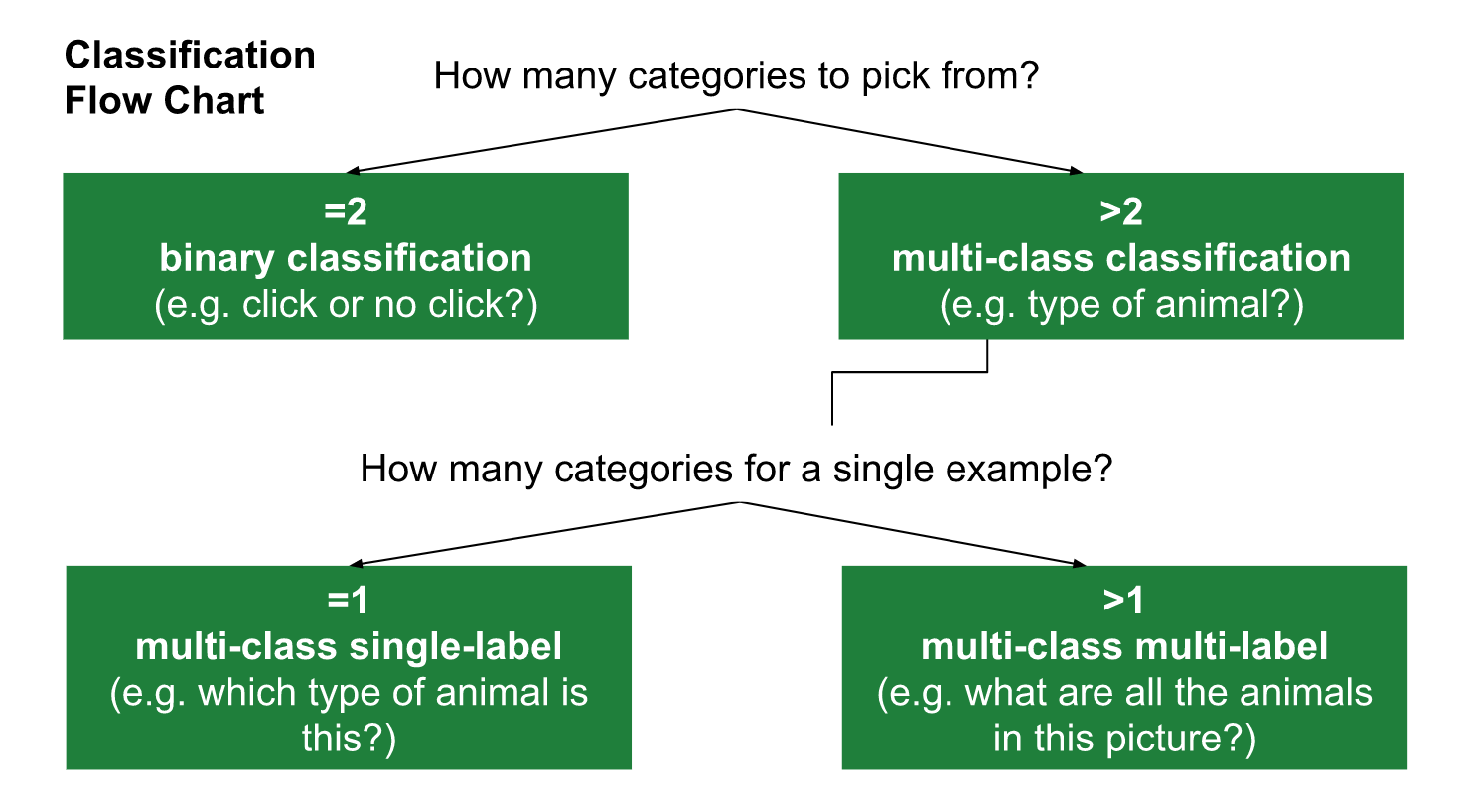
*The output will be used for..*

|  |
| --- |
|  |

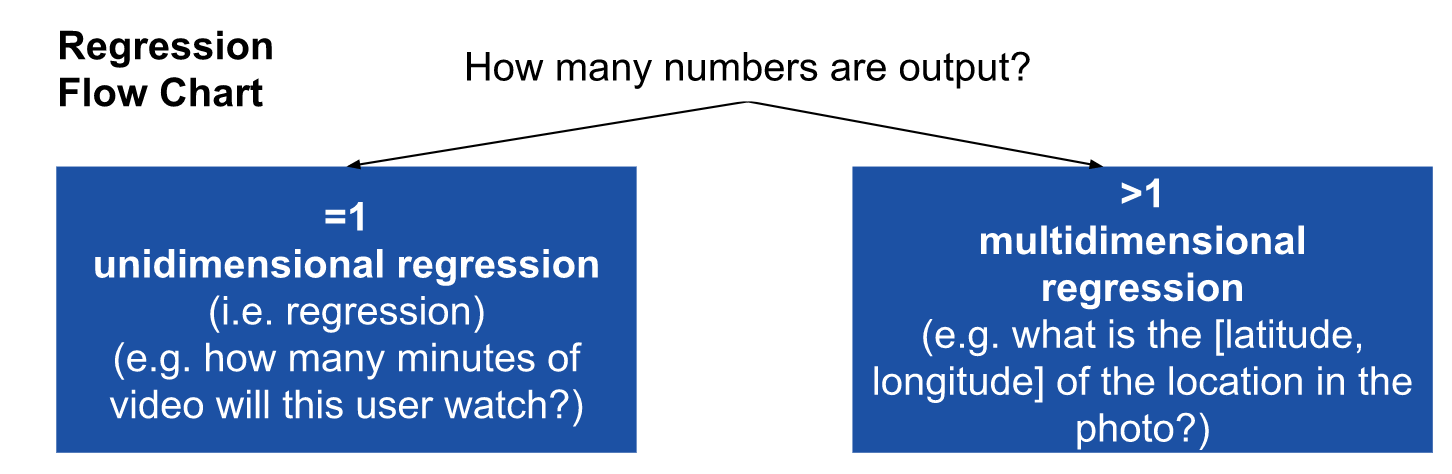
**Exercise #8: Formulating Your ML Problem (20 minutes)**

1. *Use the flowcharts below to identify your ML Problem type*

For classification problems

****

For regression problems

****

|  |  |
| --- | --- |
| **Our problem is best framed as...** |  |
| Binary classification |  |
| Unidimensional regression |  |
| Multi-class single-label classification |  |
| Multi-class multi-label classification |  |
| Multidimensional regression |  |
| Clustering (unsupervised) |  |

1. *Define your problem statement*

|  |
| --- |
| Our problem is best framed as \_\_\_\_\_\_\_, which predicts a \_\_\_\_\_\_\_\_\_\_\_\_   when \_\_\_\_\_\_\_\_\_\_\_\_\_ |

**Section 4: Get Your Data**

***Exercise #9: User-Generated Data Labelling (15 mins)***

*In the products you use every day, try to find 2 examples of user-generated data labelling.*

|  |  |  |  |
| --- | --- | --- | --- |
| Product name | What the user does | Image | How the label may be used |
|  |  |  |  |
|  |  |  |  |

***Exercise #10: Design your data for the model (15 mins)***

*Write the data you want the ML model to use to make predictions i.e. your initial features*

*For Uber Eats*

|  |  |  |  |
| --- | --- | --- | --- |
| Ave. Prep Time | Order Size | Time of Day | Preparation Time |
| mm:ss | # of items | hh:mm:ss | mm:ss |

*For Your Product*

|  |  |  |  |
| --- | --- | --- | --- |
| Feature 1 | Feature 2 | Feature 3 | Target |
|  |  |  |  |

Tips:

* For the initial model, pick easy to obtain features that you feel will deliver a reasonable initial prediction.
* Only include features that will be available at the moment the prediction is made.
* Consider where your data will come from.

**Section 5: Prepare Your Data**

***Exercise #11: Engineer New Feature Candidates (20 mins)***

*Brainstorm 2 new feature candidates for your ML model*

*Tip: Look for things that might impact your prediction, outside your initial set of features*

*For Uber Eats*

|  |  |  |  |
| --- | --- | --- | --- |
| Feature | Description | Format | *Source* |
| Food Type | To factor in cuisine type | {Soup, smoothie, sandwich, salad...} | Menu data |
| Weekend | Weekends are busier | Yes/No | Order Date |

*For Your Product*

|  |  |  |  |
| --- | --- | --- | --- |
| Feature | Description | Format | *Source* |
| Feature #1 |  |  |  |
| Feature #2 |  |  |  |

**Section 7: Deploy Your Model**

***Exercise #12 a:******Calculate Evaluation Metrics*** *(****10 mins)***

*Using the confusion matrix below, answer the following questions*

|  |  |  |
| --- | --- | --- |
|  | Predicted No | *Predicted Yes* |
| *Actual No* | 94 | 23 |
| *Actual Yes* | 24 | 100 |

*a. Number of true positives?*

*b. Number of false negatives?*

*c. Number of true negatives?*

*d. Number of false positives?*

***Exercise #12 b: Calculate Evaluation Metrics***

*Suppose we design a model to identify iphones from a video that also contain android phones. The program identifies 5 iphones in a scene containing 7 iphones and some android phones. If 3 of the identifications are correct but 2 are actually android phones…*

a. What is the precision of the model?

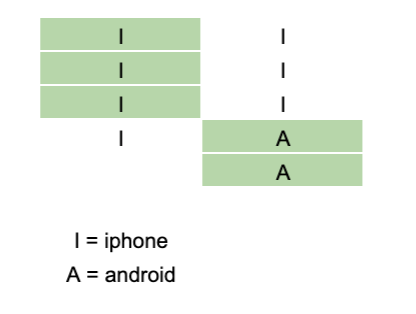
TP =

FP =   
  
Precision =

b. What is the recall of the model?

FN =

Recall =

******

***Exercise #13:******Precision, Recall or F1 Score?*** *(****10 mins)***

*In the scenarios below, would you optimize for precision, recall, or F1 Score?*

|  |  |  |  |
| --- | --- | --- | --- |
| An ML model to... | Precision | Recall | F1 Score |
| Predict whether a certain day is a good day to launch satellites based on the weather. |  |  |  |
| Detect fraudulent insurance claims. |  |  |  |
| Detect spam emails. |  |  |  |
| Detect new products that need copyright protection. |  |  |  |
| Provide a search engine. |  |  |  |