

Data Warehousing and Data Mining

Mini Project

Name:-Sanket Shivaji Thorat

Roll.no.:-33

Class:-Msc CS part 1

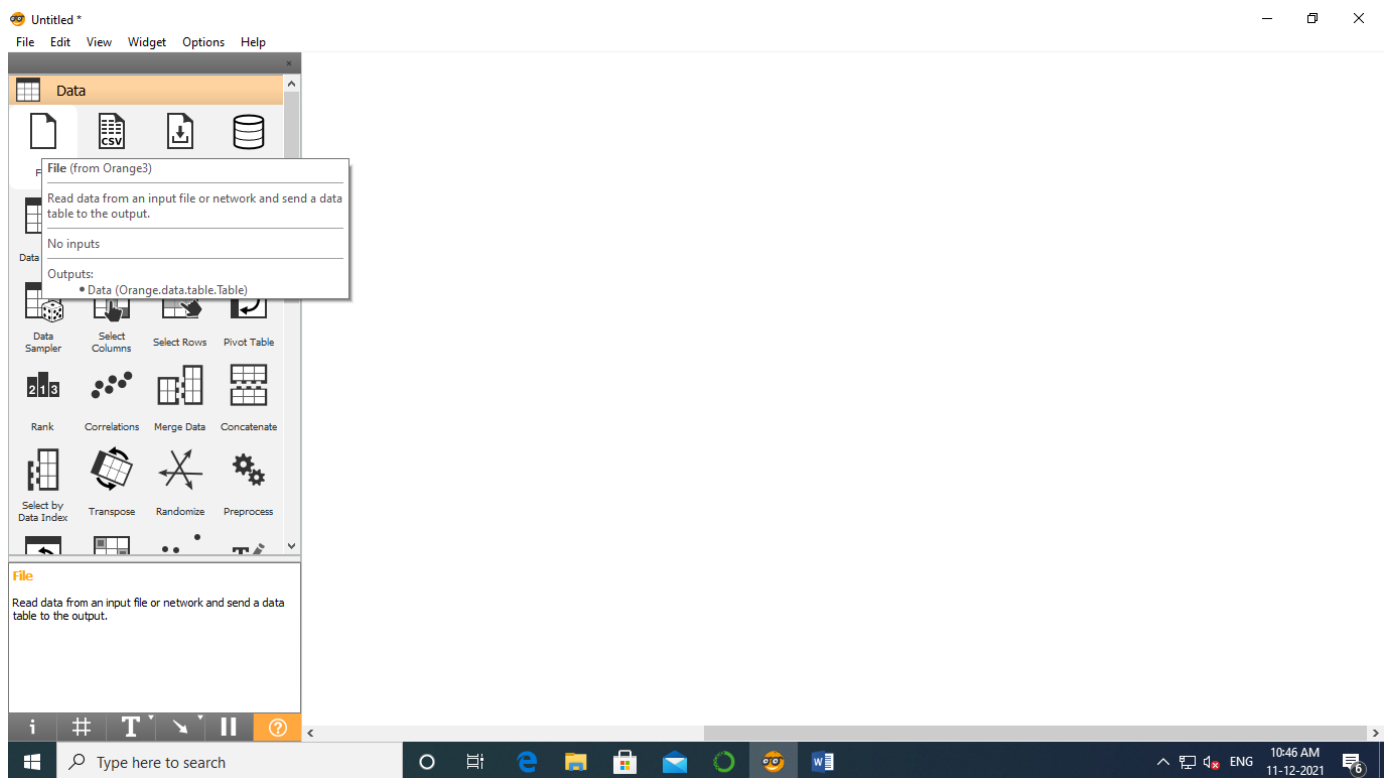
Subject:-DWDM

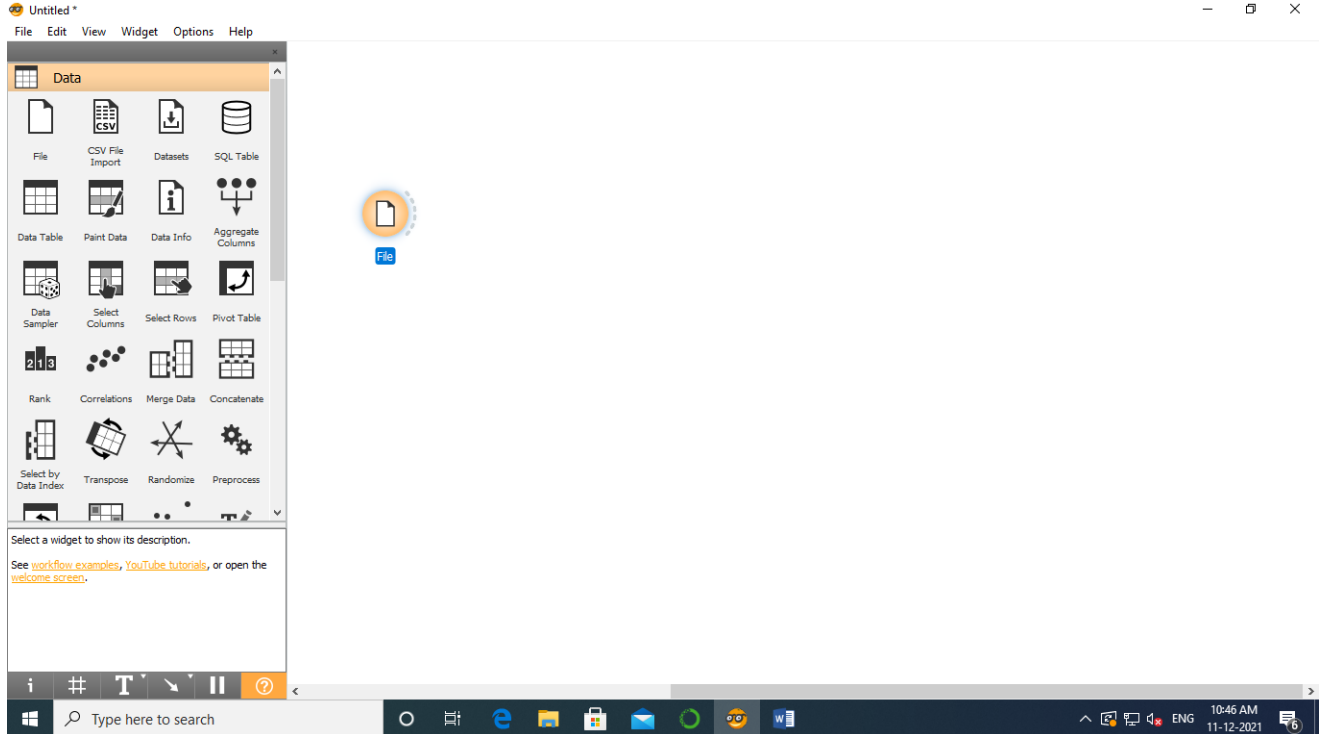
Academic Year 2021-22

Aim:- Perform hierchical clustering and classification tree on the “Iris” dataset in Orange3

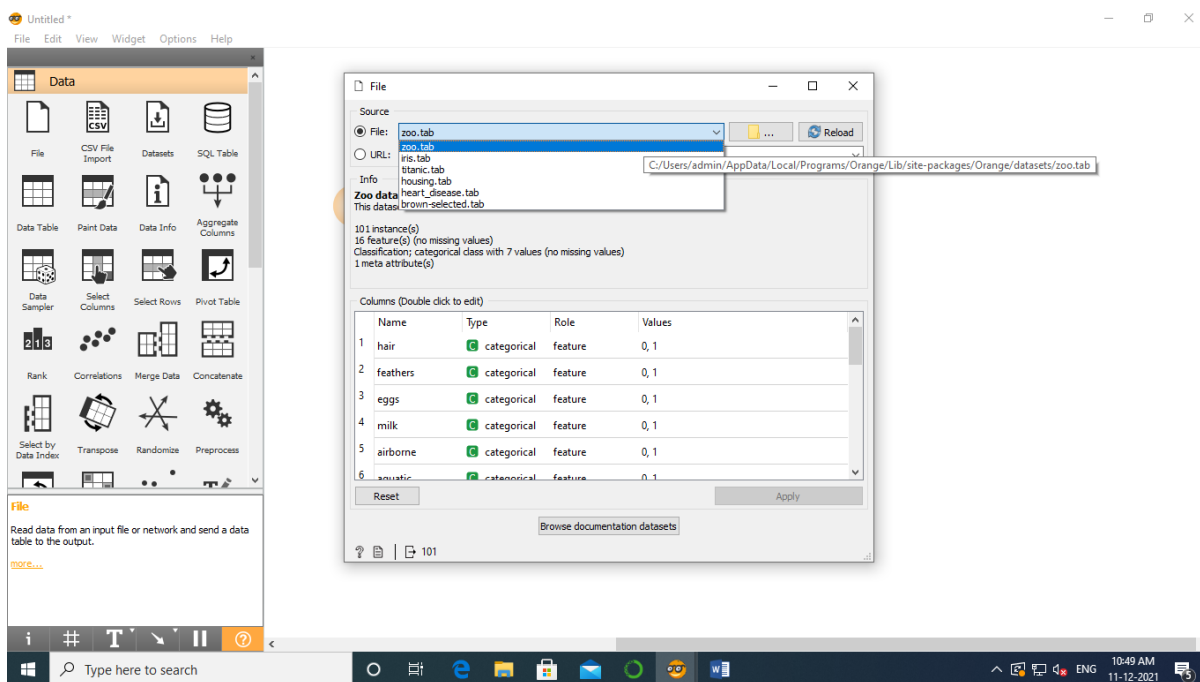
Procedure:-

Step 1: Select file and drag & drop on screen.

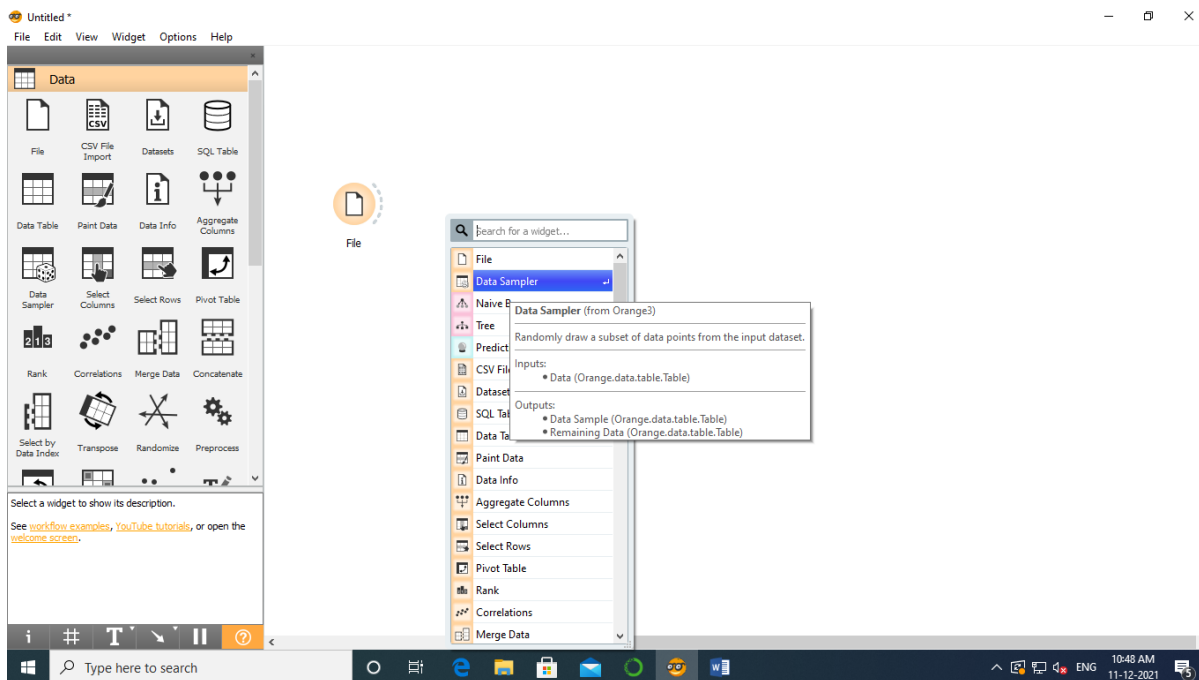




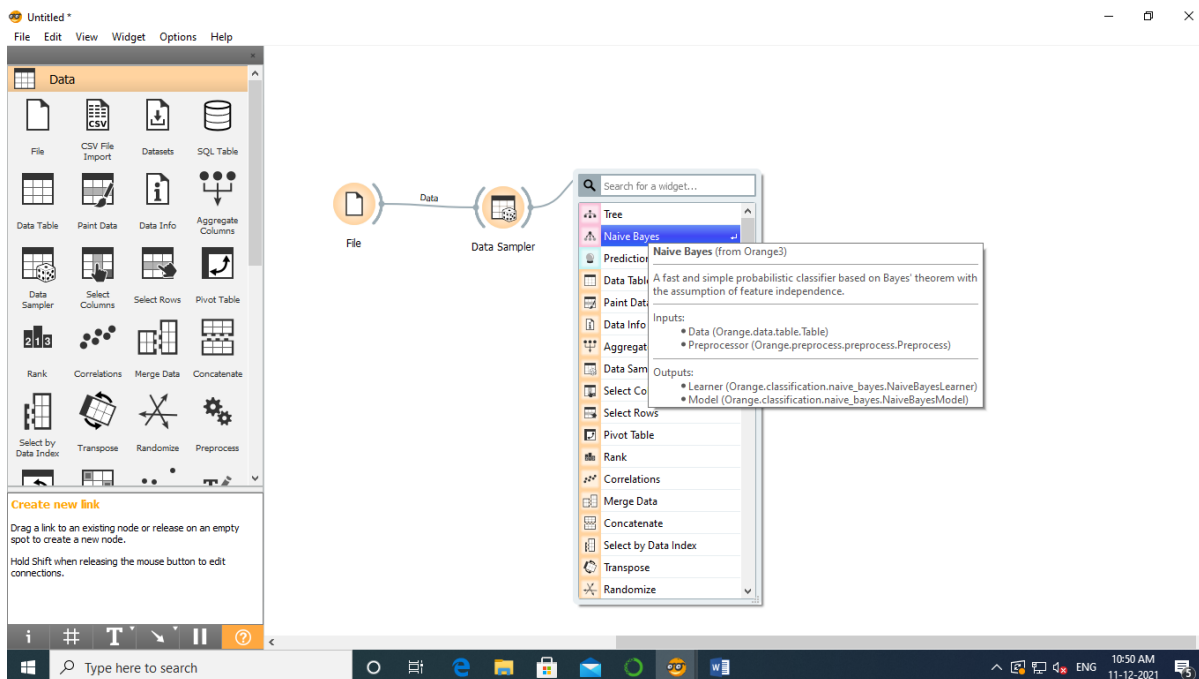
Step 2: Double click on file and select zoo.tab



Step 3: Double click on screen and select Data Sampler and connect them both



Step 4: Select Navie Bayes and connect it with Data Sampler



Step 5: Select Tree and connect it with Data Sampler

The screenshot shows the Orange3 software interface. On the left, the 'Data' widget palette is visible, containing various data processing widgets. The main workspace displays a workflow: a 'File' widget connected to a 'Data' widget, which is then connected to a 'Data Sampler' widget. The 'Data Sampler' widget has two output ports, each labeled 'Data Sample -> Data'. The top output is connected to a 'Naive Bayes' widget. The bottom output is connected to a 'Tree' widget. A context menu is open over the 'Tree' widget, showing its details: 'Tree (from Orange3)', 'A tree algorithm with forward pruning.', 'Inputs: Data (Orange.data.table.Table), Preprocessor (Orange.preprocess.preprocess.Preprocess)', and 'Outputs: Learner (Orange.modeling.tree.TreeLearner), Model (Orange.tree.TreeModel)'. The Windows taskbar at the bottom shows the time as 10:50 AM on 11-12-2021.

Step 6: Select Predication and connect it with Data Sampler

The screenshot shows the Orange3 software interface with the workflow from Step 5. The 'Tree' widget is now connected to a 'Predictions' widget. A context menu is open over the 'Predictions' widget, showing its details: 'Predictions (from Orange3)', 'Display predictions of models for an input dataset.', 'Inputs: Data (Orange.data.table.Table), Predictors (Orange.base.Model)', and 'Outputs: Predictions (Orange.data.table.Table), Evaluation Results (Orange.evaluation.testing.Results)'. The Windows taskbar at the bottom shows the time as 10:50 AM on 11-12-2021.

Step 7: Select Nomogram and connect it with Naive Bayes and double click on nomogram

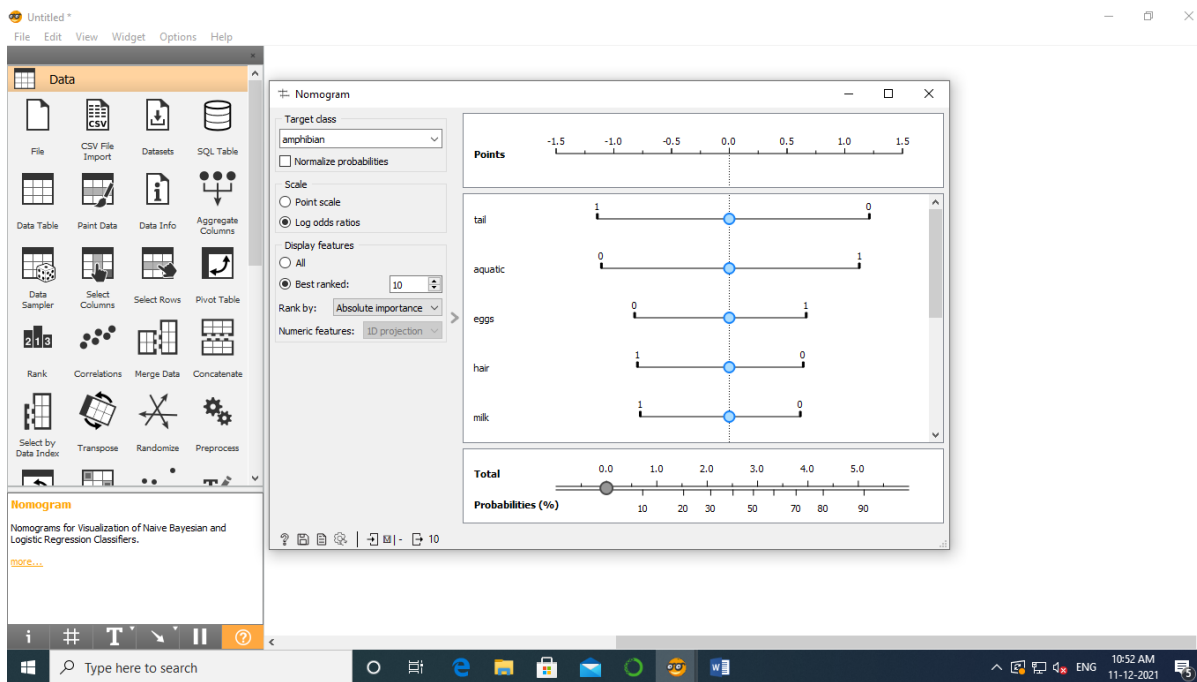
The screenshot shows the Orange3 interface with a workflow consisting of a File widget connected to a Data Sampler, which is connected to both a Naive Bayes classifier and a Tree classifier. Both classifiers are connected to a Predictions widget. A context menu is open over the Naive Bayes widget, showing the 'Nomogram' option. The 'Nomogram' widget settings are displayed in a separate window, showing the target class as 'amphibian' and the scale as 'Log odds ratios'. The 'Display features' section shows 'All' features selected, and the 'Rank by' section shows 'Absolute importance' selected. The 'Numeric features' section shows '1D projection' selected.

Workflow components:

- File
- Data Sampler
- Naive Bayes
- Tree
- Predictions

Nomogram widget settings:

- Target class: amphibian
- Normalize probabilities: ☐
- Scale: ☒ Log odds ratios
- Display features: ☒ All
- Best ranked: 10
- Rank by: Absolute importance
- Numeric features: 1D projection



Step 8: Select Navie Bayes and Tree with Predication and double click on Predication

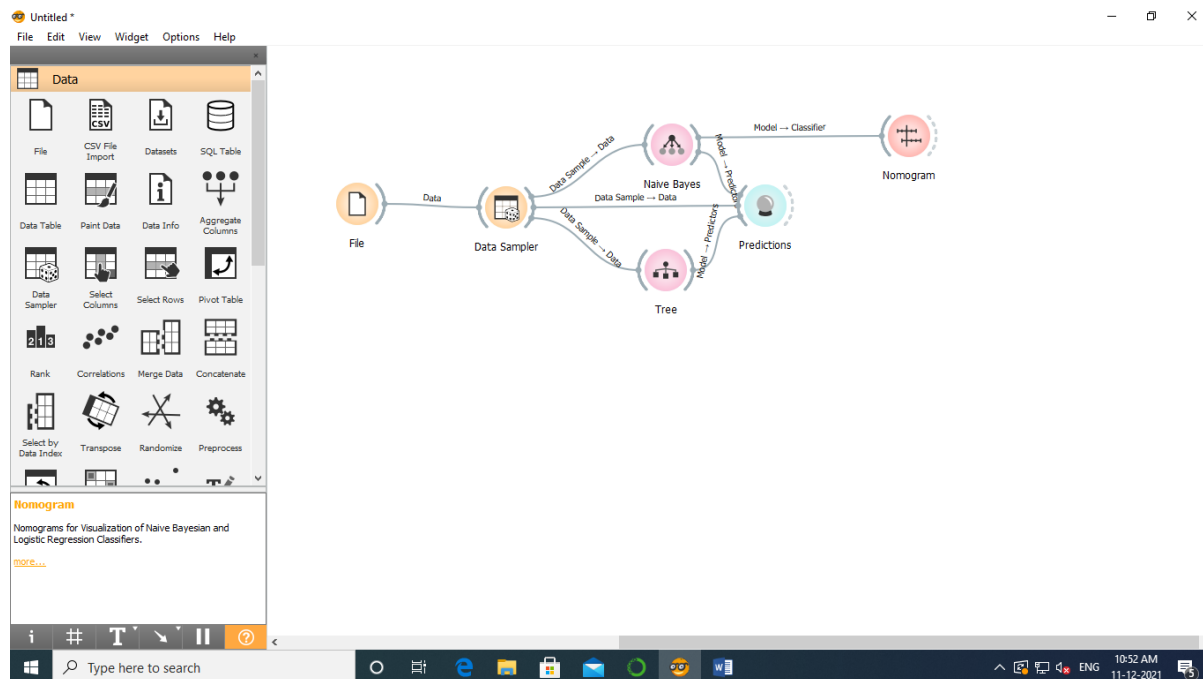


Table showing the results of the Naive Bayes and Tree models for 14 input samples. The table displays probabilities for each class (mammal, reptile, insect, fish, bird) and the predicted class. The 'Model' column indicates the model used (Naive Bayes or Tree). The 'AUC', 'CA', 'F1', 'Precision', and 'Recall' columns show the performance metrics for each model.

Model	AUC	CA	F1	Precision	Recall
Naive Bayes	1.000	0.944	0.948	0.967	0.944
Tree	0.999	0.986	0.986	0.988	0.986

Legend: p(amphibian, bird, fish, insect, invertebrate, mammal, reptile)

Step 9: Select Linear Projection and connect it with Predication

The screenshot shows the Orange3 data mining software interface. The workflow diagram includes the following components and connections:

- File** widget connected to **Data**.
- Data** widget connected to **Data Sampler**.
- Data Sampler** has two outputs, both labeled **Data Sample → Data**, which connect to **Naive Bayes** and **Predictions** respectively.
- Naive Bayes** widget connected to **Model → Classifier**.
- Model → Classifier** connected to **Nomogram**.
- Predictions** widget connected to **Linear Projection (from Orange3)**.

The **Linear Projection (from Orange3)** widget's context menu is open, displaying the following information:

- Search for a widget...**
- Linear Projection** (selected)
- Table**
- Data**
- Info**
- Aggregate Columns**
- Sampler**
- Plot Columns**
- Plot Rows**
- Pivot Table**
- Rank**
- Correlations**
- Merge Data**
- Concatenate**
- Select by Data Index**
- Transpose**
- Randomize**
- Preprocess**
- Apply Domain**

The widget description for **Linear Projection** is as follows:

- A multi-axis projection of data onto a two-dimensional plane.**
- Inputs:**
 - Data (Orange.data.table.Table)
 - Data Subset (Orange.data.table.Table)
- Outputs:**
 - Selected Data (Orange.data.table.Table)
 - Data (Orange.data.table.Table)
 - Components (Orange.data.table.Table)

This screenshot is identical to the one above, showing the same Orange3 workflow and the **Linear Projection** widget selection menu. The workflow diagram and the widget's context menu details are the same as described in the first image.

Step 10: Double click on Linear projection

