

Mid-Term Exam Review

INST414

Numpy

What is a tensor?

Slice arrays of different dimensions:

```
a3 = [[[9, 3, 5, 2, 4], [7, 6, 8, 8, 1]], [[6, 7, 7, 8, 1], [5, 9, 8, 9, 4]], [[3, 0, 3, 5, 0], [2, 3, 8, 1, 3]]]
```

```
a3[:,1,:] = ?
```

```
[[7, 6, 8, 8, 1], [5, 9, 8, 9, 4], [2, 3, 8, 1, 3]]
```

Data Wrangling

- Read a csv file: `pandas, read_table, sep="?"`
- Check the data types of the features
- Any missing values? How do you fill them?
- How do you convert categorical values ? E.g., `pandas get_dummies`
- Aggregation, e.g., `groupby, agg, sum, mean` etc

Visualization

- For which features do you use Histogram, Scatterplot, Factorplot, Pairplot ?
- Covariance matrix of multivariate Gaussian, e.g., meaning of diagonal and off-diagonal entries

Probability Distributions

- Scipy.stats and Numpy for normal, bernoulli, and binomial distributions
- Entropy of a random variable, e.g., use scipy.stats to calculate entropy
- Bootstrapping, e.g., estimate the confidence interval of a parameter

Linear and Logistic Regression

- Supervised vs Unsupervised; Regression vs Classification
- Loss (error) functions (aka cost functions), e.g., MAE, RMSE, cross-entropy
- Confusion matrix, e.g., accuracy and fairness, F1 score