\documentclass{article} % For LaTeX2e

\usepackage{nips15submit\_e,times}

\usepackage{hyperref}

\usepackage{url}

\usepackage{graphicx}

\usepackage{subcaption}

\graphicspath{ {figures/} }

\usepackage{amsmath}

\usepackage{multirow}

\usepackage{biblatex}

\addbibresource{sample.bib}

\usepackage[justification=centering]{subfig}

\usepackage{subcaption}

%\documentstyle[nips14submit\_09,times,art10]{article} % For LaTeX 2.09

\title{Airline Fuel Burn Predictions and Prediction Intervals}

\author{

Lei Kang \\

Department of Civil Engineering\\

University of California, Berkeley\\

Berkeley, CA 94720 \\

\texttt{kang119@berkeley.edu } \\

\And

Madeleine Sheehan \\

Department of Civil Engineering\\

University of California, Berkeley\\

Berkeley, CA 94720 \\

\texttt{m.sheehan@berkeley.edu} \\

}

% The \author macro works with any number of authors. There are two commands

% used to separate the names and addresses of multiple authors: \And and \AND.

%

% Using \And between authors leaves it to \LaTeX{} to determine where to break

% the lines. Using \AND forces a linebreak at that point. So, if \LaTeX{}

% puts 3 of 4 authors names on the first line, and the last on the second

% line, try using \AND instead of \And before the third author name.

\newcommand{\fix}{\marginpar{FIX}}

\newcommand{\new}{\marginpar{NEW}}

%\nipsfinalcopy % Uncomment for camera-ready version

\begin{document}

\maketitle

\begin{abstract}

Airlines are moving aggressively to reduce fuel consumption. Obtaining accurate trip fuel prediction is a critical piece in achieving that goal. In this study, we propose to use Lasso selection to combine different learners into a super learner which can reduce prediction error by 50\% compared to our study airline’s flight planning system. In order to gauge the uncertainty associated with individual predictions, we group flights into different clusters having similar model errors using stability-based K-means and Gaussian mixture clustering techniques. The prediction interval is constructed for each cluster on the basis of empirical distributions of the prediction errors associated with all observations belonging to the cluster. The clustering idea allows us to achieve similar prediction interval coverage probability with tighter interval limits. %edit this

\end{abstract}

\section{Introduction/Objective}

Accurate airplane fuel burn predictions are crucial in the aviation industry. The required trip fuel is calculated by airline’s Flight Planning System (FPS) based on factors including the route selection, altitude structure, winds and other weather conditions forecast, anticipated traffic delays, and aircraft performance. However, the FPS sometimes performs unsatisfactory jobs in planning trip fuel. If planned trip fuel is higher than actual trip fuel, then it simply indicates fuel overloading and a flight will burn extra fuel to carry those unnecessary loaded fuel (more fuel cost). On the other hand, if trip fuel is under-estimated, then a flight might run into fuel emergency issue (safety concern). Therefore, it is critical to have accurate and reliable trip fuel predictions.

In practice airlines carry extra fuel to avoid the latter (under-prediction) case. However, carrying extra fuel can prove costly and harmful to the environment as airlines burn more fuel when carrying this extra fuel weight. Based on the study of a major U.S-based airline \cite{ryerson}, reducing unnecessary fuel loading could result in airline fuel savings on the order of \$400 million per year. Some of this extra fuel burn can be avoided by more accurate fuel predictions and a better understanding of the uncertainty in those predictions.

The goals of this paper are three-fold. First, we use several machine learning models to improve predictions on airline fuel burn. Second, we develop a super-learner that combines the results from the various machine learning methods. Third, we estimate the uncertainty in our model by partitioning the input space into clusters where the points in each cluster exhibit a similar distributions of model error. We then construct a prediction interval rather than a single point prediction for all points in the validation set.

\section{About the data set}

A major U.S.-based airline provided detailed flight-level information from its domestic flights between April 2012 and July 2013. Table 1 presents a brief summary of predictors we will use in building prediction models for actual fuel burn (in lbs). There are a total of 465,404 domestic flights in the airline data (with eight major aircraft types) and this data set contains basic flight operation characteristics, for instance aircraft type, origin and destination (OD) airports, planned airborne time, schedule and actual departure and arrival information, and actual trip fuel burn quantities (in pounds). The study airline also has its own Flight Planning System (FPS) which predicts trip fuel burn for each flight. There is often large discrepancy between planned and actual trip fuel burn which leaves great potential to improve fuel burn prediction.

The airline data set does not include forecast or actual weather information, the weather data instead comes from the National Oceanic and Atmospheric Administration (NOAA) database. The weather information used recreates the information available to flight dispatchers during the time of flight fuel planning (2 hours prior to scheduled departure time).

However, we do not have forested en-route weather or air traffic information which presumably plays a critical role in predicting fuel burn. Historical flight level information from the Aviation System Performance Metrics (ASPM) serves as a proxy - for a given flight in the airline data, we look at historical flights with same OD pair, scheduled departure hour, and month that occurred in the previous year. Then we can calculate say mean, standard deviation of airborne time, and different quantiles of airborne time based on these historical similar flights. By assuming stable weather patterns for a given OD, month, and hour of day, this historical airborne time information will provide good approximation in terms of possible weather conditions for a current flight. The deviation of historical airborne times and the corresponding flight plan airborne times serves as another measure of flight time variability.

The second proxy we will use is the Airline's FPS planned trip fuel burn. FPS takes weather and traffic forest information from various resources as inputs to estimate planned trip fuel. Though we have no access to those original data source, we can still use FPS planned trip fuel as proxy for en-route conditions and build our prediction upon the FPS output.

The data set is randomly split into three categories: 372,323 observations as training set, 45,000 observations as holdout set to construct prediction intervals, and 48,081 observations as test set.

\begin{table}[h!]

\centering

\begin{tabular}{|c|c|c|c|}

\hline

Category & Variable & Note & Source \\

\hline

\multirow{5}{\*}{Weather} & Forecasted ceiling condition at destination airport &Forecasted& \multirow{4}{\*} {NOAA}\\

\cline{2-2}

&Forecasted visibility condition at destination &weather at&\\

\cline{2-2}

&Forecasted thunderstorm condition at destination & time of flight&\\

\cline{2-2}

&Forecasted snow condition at destination &planning&\\

\cline{2-4}

&\multirow{2}{\*}{Number of alternate airports}&Indication of&\multirow{2}{\*}{Airline}\\

&&weather&\\

\hline

Historical & Distribution of historical airborne time & Based on historical & \multirow{4}{\*}{ASPM}\\

Traffic & (e.g. mean, maximum, and percentiles) & flights with similar & \\

\cline{2-2}

Information & Deviation of the airborne time from flight& OD, month,&\\

&plan for historical flights & and departure time&\\

\hline

\multirow{2}{\*}{Airport}&Coordinates of origin and destination airports&\multirow{2}{\*}{ -- }&\multirow{2}{\*}{Airline}\\

& great circle distance between OD && \\

\hline

\multirow{3}{\*}{Flight}& Aircraft type, fuel capacity, tanker fuel, & Flight level & \multirow{3}{\*}{Airline}\\

&alternate fuel, planned trip time, planned &information&\\

& trip fuel, departure hour and month&&\\

\hline

\end{tabular}

\caption{Model input variables}

\label{table:1}

\end{table}

\section{Methodology}

\subsection{Predictions}

The stacking or super learner method is an ensemble learning technique used by !!NAMES!! \cite{wolpert} \cite{breiman} \cite{leblanc} \cite{tibshirani} \cite{vanderlaan}. It seeks optimal linear combinations of base learners to improve prediction accuracy. Suppose we have selected $K$ learners $f\_1(\textbf{x})...f\_K(\textbf{x})$ to reproduce outcome variable $y$, on a training set, $S= \{(y\_i,\textbf{x}\_i)=1,...,N$, then instead of following the winner-takes-all strategy and selecting a single learner based on lowest cross validation error, we can stack those learners to form a new design matrix and combine them in the following form:

\begin{equation} \label{eq:SL1}

SL(\textbf{x}) = \sum\_{k=1}^{K}a\_kf\_k(\textbf{x})

\end{equation}

where SL stands for super learner and are the weights associated with each learner. Given the training set $\{y\_i, \textbf{x}\_i), i=1...N\}$ we choose $\{a\_k\}$ to minimize:

\begin{equation} \label{eq:SL2}

\sum\_i(y\_i-\sum\_{k=1}^Ka\_kf\_k(\textbf{x}\_i))^2

\end{equation}

To some extent, we can leverage the strength of each individual learner and achieve better prediction performance. However, there are two potential issues if we obtain $\{a\_k\}$ using ordinary least squares on equation (2). The first issue is overfitting, as Brieman pointed out \cite{breiman}, if the $f\_k(\textbf{x})$ is constructed using the training set $S$, and the $\{a\_k\}$ are also selected by minimizing squared error over $S$, then the resulting $\{a\_k\}$ will overfit the data. As remedy, Wolpert suggest to use leave-one-out CV to generate $f\_k^{LOOCV}(\textbf{x})$ and Breiman proposed to use k-fold CV to generate $f\_k^{CV}$ \cite{wolpert} \cite{breiman}. The ${a\_k}$ are then chosen to minimize

\begin{equation} \label{eq:SL3}

\sum\_i(y\_i-\sum\_{k=1}^Ka\_kf\_k^{CV}(\textbf{x}\_i))^2

\end{equation}

The second issue is how to estimate $\{a\_k\}$. Breiman found that when the coefficients in (\ref{eq:SL3}) were constrained to be nonnegative and $\sum\_k a\_k = 1$ the super learner showed better prediction error than any individual $f\_k(\textbf{x})$ \cite{breiman}. However, the solution of equation (\ref{eq:SL3}) obtained using nonnegative least squares might be too restrictive in finding feasible solutions \cite{lawson}.

%%Edit this!!!

We will discuss in detail the numeric findings using non-negative least squares in the later section.

In this study, we propose to use LASSO to estimate constrained $\{a\_k\}$ in a more data adaptive manner \cite{tibshirani}. Since all the base learners are predicting the same y, their outputs will be highly correlated. Thus, LASSO will help shrink the weights of duplicated learners to zero and help reduce overfitting. Now the minimization problem becomes:

\begin{equation} \label{eq:Lasso\_SL}

\min\_a \sum\_i(y\_i-\sum\_{k=1}^K a\_k f\_k^{CV}(\textbf{x}\_i))^2

\end{equation}

\begin{equation}

s.t. ||a||\_1\le t

\end{equation}

or equivalently,

\begin{equation} \label{eq:Lasso\_SL2}

\min\_a \sum\_i(y\_i-\sum\_{k=1}^K a\_k f\_k^{CV}(\textbf{x}\_i))^2 + \lambda ||a||\_1

\end{equation}

The regularization parameter is selected using k -fold cross validation.

\subsection{Prediction intervals}

Several methods have been proposed for computing prediction intervals for neural networks and other non-linear models. Many rely on derivation of the error variance $s^2$, a procedure that is not trivial for non-linear models. To do this, some have made bold assumptions about constant variance in the input space and assume independence between the predictions, $\hat{\theta}$, and the residuals $\epsilon$ (Chryssoloiuris, Lee, and Ramsey). Others try to estimate $s^2$ with re-sampling techniques. These methods rely on ensemble learning to estimate model variance and then use an auxiliary model to estimate the target noise. These models are often very computationally intensive, and typically ignore model bias.

Instead we followed the procedure outlined in (Shrestha and Solomatine). The main idea is that historical residuals are the best indicators of future residuals. In other words the observed differences between predicted values and observed data give the best indication of future prediction errors. The prediction intervals will therefore computed from empirical distributions of residuals.

The steps to compute and validate the prediction intervals are outlined here.

First, partition the data into clusters with similar model errors. If the data can be cleanly partitioned into distinct clusters, using k-means for example, the computation of the empirical prediction intervals is straight forward. The limits of the cluster's prediction interval are set to encompass $1-\alpha\%$ of the residuals in the cluster. The bounds of the interval are set to the $\alpha/2$ and $1-\alpha/2$ percentiles of the residuals, as shown in \ref{fig:cdf}. The prediction interval for each data point in the cluster is equal to the prediction interval of the cluster.

\begin{figure}[h]

\centering

\includegraphics[width =.3\textwidth]{CDF}

\caption{Constructing prediction intervals}

\label{fig:cdf}

\end{figure}

%may want to add details about determining cluster prediction interval if fuzzy k-means/gaussian

If the clusters are not clearly defined, (if for example a fuzzy k-means or gaussian mixture model is used to partition the data) then the prediction interval is a weighted combination of the cluster prediction intervals - weighted by percent membership in each cluster.

PI validation is not as straight forward as validation of the point predictions. Unlike the point estimation, the target prediction interval is unknown, so it can be difficult to validate. However, one way of checking the validity of the prediction interval is to test what percentage of the time prediction interval includes the observed value on a validation set. By definition the prediction interval should include the target value $(1-\alpha)\%$ of the time. The percent of observed target values that fall within the validation set can be computed for each cluster. This number is refered to as the prediction interval coverage probability (PICP).

% \section{Results}

% \subsection{Prediction results}

% \begin{table}[h!]

% \centering

% \begin{tabular}{|c|c|c|c|c|c|}

% \hline

% Airline & \multicolumn{4}{|c|}{ML Algorithms}& Super learner \\

% \hline

% Flight & Random & Ridge & Lasso & GBM & SL Lasso \\

% Planning System & Forests

% & & & &

% (0.125\*Ridge + 0.875\*RF)\\

% \hline

% 950,431 & 473,915 & 516,337 & 516,306 & 491,484 & 467,071\\

% \hline

% \end{tabular}

% \caption{Mean-squared error on test sets}

% \label{table:results}

% \end{table}

\section{Results}

\subsection{Prediction results}

\begin{table}[h!]

\centering

\setlength\tabcolsep{4.5pt}

\begin{tabular}{|c|c|c|c|c|c|c|}

\hline

& 5-fold CV & Weights of& Weights of&Test risk&Test risk&Test risk\\

Base & Risk (MSE) & Super & original & (MSE) using & using & (MSE)\\

Learners & on training & Learner with & Super & Model & Super & Airline FPS \\

& set (lbs$^2$) & LASSO & Learner & Prediction & Learner & estimates \\

&&selection&&(lbs$^2$)&(lbs$^2$)&(lbs$^2$)\\

\hline

Random & \multirow{3}{\*}{467,046} & \multirow{3}{\*}{0.875} & \multirow{3}{\*}{1}& \multirow{3}{\*}{473,915} &\multirow{8}{\*}{467,071} &\multirow{8}{\*}{950,431} \\

Forest &&&&&& \\

(1000 trees) &&&&&& \\

\cline{1-5}

Gradient & \multirow{3}{\*}{470,603} & \multirow{3}{\*}{0} & \multirow{3}{\*}{0}& \multirow{3}{\*}{491,484} & & \\

Boosting &&&&&& \\

tree &&&&&& \\

\cline{1-5}

Ridge & 518,188 & 0.125 & 0 & 516,337&& \\

\cline{1-5}

Lasso & 518,193 & 0 & 0 & 516,306&& \\

\hline

\end{tabular}

\caption{Summary of prediction results.}

\label{table:results}

\end{table}

Notes:

1. Tuned parameters for random forests: max depth 25. Planned trip fuel burn is found to be the most important factor. Way more important than others.

2. For gradient boosting, number of iterations is decided based on 20\% validation set, not CV since it takes very long time to run. The selected learning rate is 0.01, minimum number of observations in a terminal node is set to be 10 and maximum number of iterations is 14220.

Superlearner table,

Prediction accuracy vs. flight plan

\subsection{Prediction interval results}

\subsubsection{Clustering}

After obtaining individual predictions, the next task is to construct prediction intervals associated with each prediction. The input dataset can be partitioned into homogeneous groups that are likely to share similar prediction errors. Prediction intervals are constructed for each cluster on the basis of the empirical distribution of errors.

For the fuel burn predictions we perform two levels of clustering, first based on the coordinates of the origin and destination airports - this allows us to capture directional effects (e.g. wind and regional congestion effects), then based on ((other flight properties - UPDATE THIS)). We will refer to the first layer of clustering OD cluster (for origin-destination clusters), and the second sub-clusters.

The OD clustering is achieved with K-means clustering. One challenge of using K-means is in deciding the optimal number of clusters. We implement a stability-based K-means clustering algorithm proposed by Ben-hur et al. (2002). The basic idea of stability-based clustering is that “true” clusters should be insensitive to data perturbation. Thus, if we randomly select two subsamples of original data (for instance each accounts for 60\% of original sample size) and perform K-means on each subsample, we will be able to assess how stable clustering assignments are for the same observations that are shared by two subsamples. Concretely, we can express the clustering labels L of data X using a matrix C with components:

\begin{align}

C\_{ij}=

\begin{cases}

1,& \text{if } x\_i \text{ and } x\_j \text{ belong to the same cluster and } i \ne j\\

0,& \text{ otherwise}

\end{cases}

\end{align}

Let $L\_1$ corresponds to the labels from the first sub-sample of data (accounting for 60\% of the original data) with matrix representation $C^1$, and $L\_2$ correspond to the second subsample with matrix representation $C^2$. The dot product $<C^1,C^2> = \sum\_{i,j}C\_{ij}^1C\_{ij}^2$ computes the number of pairs of points clustered together. Then we can define two similarity measures. The first one is correlation similarity measure:

\begin{equation}

cor(L\_1,L\_2)=\frac{<C^1,C^2>}{\sqrt{<C^1,C^1><C^2,C^2>}}

\end{equation}

The second one is called Jaccard coefficient defined as:

\begin{equation}

jac(L\_1,L\_2) = \frac{<C^1,C^2>}{<C^1,C^1>+<C^2,C^2>-<C^1,C^2>}

\end{equation}

Both metrics are based on how well entries of the two matrices agree and the values of metrics vary between 0 and 1. Higher metric value indicates more stable clustering assignment. For number of clusters ranging from 2 to 15, we perform the above mentioned resample and K-means algorithm 100 times and display the distribution of corresponding similarity measures. As shown in Figure \ref{fig:corr6} based on the coordinates of origin airports, the most stable number of clusters is 3 since most of its measures are close to 1.

\begin{figure}[h]

\begin{center}

\begin{subfigure}{0.48\textwidth}

\includegraphics[width=\textwidth]{correlation6}

\caption{Correlation similarity measure}

\label{fig:corr}

\end{subfigure}

\begin{subfigure}{0.48\textwidth}

\includegraphics[width=\textwidth]{jac6}

\caption{Jaccard Similarity Measure}

\label{fig:jac}

\end{subfigure}

\end{center}

\caption{Cumulative density of similarity metric for k=2 to 15 clusters}

\label{fig:corr6}

\end{figure}

\begin{figure}[h]

\begin{center}

\begin{subfigure}{0.48\textwidth}

\includegraphics[width=\textwidth]{od\_cluster1}

\caption{OD Cluster 1}

\label{fig:spatial1}

\end{subfigure}

\begin{subfigure}{0.48\textwidth}

\includegraphics[width=\textwidth]{od\_cluster2}

\caption{OD Cluster 2}

\label{fig:spatial2}

\end{subfigure}

\begin{subfigure}{0.48\textwidth}

\includegraphics[width=\textwidth]{od\_cluster3}

\caption{OD Cluster 3}

\label{fig:spatial3}

\end{subfigure}

\label{fig:spatial\_clust}

\end{center}

\caption{Spatial clustering of flight origin and destination airports}

\end{figure}

After determining optimal number of clusters, we run K-means 20 times and select the cluster assignment corresponds to the lowest total within cluster sum of squares to reduce the influence of random initialization. The resulting OD clusters are shown in Figure \label{fig:spatial\_clust}. The three clusters generally correspond to east bound, west bound, and eastern traffic. This layer 1 clustering allows us to separate flights into groups with similar traffic patterns.

In order to achieve more accurate PI construction, within each layer 1 cluster, we perform another layer of clustering based on individual flight characteristics: historical mean, 75th percentile, and maximum airborne time, the distribution of historical flights’ deviation from flight plan (measured by mean, 75th percentile, and maximum values), planned trip time and trip fuel burn, and forecasted weather conditions at the destination airports. For layer 2 clustering, we apply Gaussian mixture models by assuming our input features following multivariate Gaussian distributions with full covariance structure. Compared to K-means, Gaussian mixture clustering is a soft-version clustering technique as it estimates the probabilities of each flight belonging to a cluster. This soft version clustering allows us to utilize information from all sub-clusters in constructing PI. %%MAYBE cut out some description of properties.

The optimal number of Gaussian distributions (sub-clusters) per OD cluster is determined based on Bayesian Information Criterion (BIC). BIC is an indicator of model fit where lower BICs indicate a better fit. Shown in the BIC plot, as we increase number of Gaussians, BIC decreases. However, after certain numbers, the marginal decrease becomes negligible. Therefore we stop at the position where there is no subsequent significant improvement in model fit and pick the corresponding number as our optimal number of clusters. One reason of adopting this early stopping is that we generally prefer a simple model to complicated models. In the end, we find 5 sub-clusters for OD cluster 1 and OD cluster 2, and 3 sub-clusters for OD cluster 3.

To get a sense of how Gaussian mixture is doing, we perform PCA on flights from each layer 1 cluster to reduce input features into 3 principal components. It is observed that the first 3 principal components have already explained more than 65\% variance across all three cases. By examining factor loadings on each component, we find that the 1st component generally corresponds to travel time, 2nd component corresponds to flight plan deviation, and the 3rd component corresponds to weather conditions. Here we visualize sub-clusters for each layer 1 cluster by projecting flights onto their first three principal components. It is interesting to notice that most flights can be distinguished well by the first three principal components. For instance, in OD cluster 1, we can easily identify short haul sub-clusters which also justify the good performance of Gaussian mixture clustering.

\begin{figure}

\centering

\includegraphics[width =\textwidth]{Gaussian\_mixture2}

\begin{minipage}[t]{.32\linewidth}

\centering

\subcaption{OD Cluster 1}\label{bic1}

\end{minipage}%

\begin{minipage}[t]{.32\linewidth}

\centering

\subcaption{OD Cluster 2}\label{bic2}

\end{minipage}

\begin{minipage}[t]{.32\linewidth}

\centering

\subcaption{OD Cluster 3}\label{bic2}

\end{minipage}

\caption{Bayesian Information Criterion (BIC) to select number of subclusters}

\label{fig:my\_label}

\end{figure}

\subsubsection{Prediction Intervals}

Decision trees and random forests are known for overfitting. Model overfitting is the case where the training error is smaller than the prediction error. When this occurs, prediction intervals that are constructed from the residuals of the training set will be too narrow. Such is the case for the random forest fuel burn predictor. As shown in Figure \ref{fig:picp} a prediction interval that covers 99\% of the residuals in the training set only has an 83\% prediction interval coverage probability on the test set. The $\ell\_1$ norm penalty of the lasso super learner model somewhat combats the overfitting of the random forest, but still the prediction interval that covers 99\% of the residuals in the training set only covers 86\% of the test residuals. In contrast, a prediction interval constructed from a separate holdout set that was not used to train the models maintains a 99\% coverage probability on the test set. From this point forward we will use only the prediction intervals generated from the holdout set.

\begin{figure}[h]

\begin{center}

\begin{subfigure}{0.32\textwidth}

\includegraphics[width=\textwidth]{Cluster\_1\_PICP}

\caption{Cluster 1}

\label{fig:picp1}

\end{subfigure}

\begin{subfigure}{0.32\textwidth}

\includegraphics[width=\textwidth]{Cluster\_2\_PICP}

\caption{Cluster 2}

\label{fig:picp2}

\end{subfigure}

\begin{subfigure}{0.32\textwidth}

\includegraphics[width=\textwidth]{Cluster\_3\_PICP}

\caption{Cluster 3}

\label{fig:picp3}

\end{subfigure}

\end{center}

\caption{Prediction interval coverage probabilities.}

\label{fig:picp}

\end{figure}

\begin{figure}[h]

\begin{center}

\begin{subfigure}{0.32\textwidth}

\includegraphics[width=\textwidth]{Cluster\_1\_prediction\_interval\_limits}

\caption{OD Cluster 1}

\label{fig:picp1}

\end{subfigure}

\begin{subfigure}{0.32\textwidth}

\includegraphics[width=\textwidth]{Cluster\_2\_prediction\_interval\_limits}

\caption{OD Cluster 2}

\label{fig:picp2}

\end{subfigure}

\begin{subfigure}{0.32\textwidth}

\includegraphics[width=\textwidth]{Cluster\_3\_prediction\_interval\_limits}

\caption{OD Cluster 3}

\label{fig:picp3}

\end{subfigure}

\end{center}

\caption{Prediction interval limits.}

\end{figure}

Report CIs (at least for smallest and largest CI group),

\subsubsection{Validation}

Validate on test set

\section{Conclusion and Future Work}

%Interesting extension would be to model the residuals as a function of the inputs with an artificial NN or RF or other means.

\newpage

\medskip

\printbibliography

--------------------------------------------------------

% NIPS requires electronic submissions. The electronic submission site is

% \begin{center}

% \url{http://papers.nips.cc}

% \end{center}

% Please read carefully the

% instructions below, and follow them faithfully.

% \subsection{Style}

% Papers to be submitted to NIPS 2015 must be prepared according to the

% instructions presented here. Papers may be only up to eight pages long,

% including figures. Since 2009 an additional ninth page \textit{containing only

% cited references} is allowed. Papers that exceed nine pages will not be

% reviewed, or in any other way considered for presentation at the conference.

% %This is a strict upper bound.

% Please note that this year we have introduced automatic line number generation

% into the style file (for \LaTeXe and Word versions). This is to help reviewers

% refer to specific lines of the paper when they make their comments. Please do

% NOT refer to these line numbers in your paper as they will be removed from the

% style file for the final version of accepted papers.

% The margins in 2015 are the same as since 2007, which allow for $\approx 15\%$

% more words in the paper compared to earlier years. We are also again using

% double-blind reviewing. Both of these require the use of new style files.

% Authors are required to use the NIPS \LaTeX{} style files obtainable at the

% NIPS website as indicated below. Please make sure you use the current files and

% not previous versions. Tweaking the style files may be grounds for rejection.

% %% \subsection{Double-blind reviewing}

% %% This year we are doing double-blind reviewing: the reviewers will not know

% %% who the authors of the paper are. For submission, the NIPS style file will

% %% automatically anonymize the author list at the beginning of the paper.

% %% Please write your paper in such a way to preserve anonymity. Refer to

% %% previous work by the author(s) in the third person, rather than first

% %% person. Do not provide Web links to supporting material at an identifiable

% %% web site.

% %%\subsection{Electronic submission}

% %%

% %% \textbf{THE SUBMISSION DEADLINE IS June 5, 2015. SUBMISSIONS MUST BE LOGGED BY

% %% 23:00, June 5, 2015, UNIVERSAL TIME}

% %% You must enter your submission in the electronic submission form available at

% %% the NIPS website listed above. You will be asked to enter paper title, name of

% %% all authors, keyword(s), and data about the contact

% %% author (name, full address, telephone, fax, and email). You will need to

% %% upload an electronic (postscript or pdf) version of your paper.

% %% You can upload more than one version of your paper, until the

% %% submission deadline. We strongly recommended uploading your paper in

% %% advance of the deadline, so you can avoid last-minute server congestion.

% %%

% %% Note that your submission is only valid if you get an e-mail

% %% confirmation from the server. If you do not get such an e-mail, please

% %% try uploading again.

% \subsection{Retrieval of style files}

% The style files for NIPS and other conference information are available on the World Wide Web at

% \begin{center}

% \url{http://www.nips.cc/}

% \end{center}

% The file \verb+nips2015.pdf+ contains these

% instructions and illustrates the

% various formatting requirements your NIPS paper must satisfy. \LaTeX{}

% users can choose between two style files:

% \verb+nips15submit\_09.sty+ (to be used with \LaTeX{} version 2.09) and

% \verb+nips15submit\_e.sty+ (to be used with \LaTeX{}2e). The file

% \verb+nips2015.tex+ may be used as a ``shell'' for writing your paper. All you

% have to do is replace the author, title, abstract, and text of the paper with

% your own. The file

% \verb+nips2015.rtf+ is provided as a shell for MS Word users.

% The formatting instructions contained in these style files are summarized in

% sections \ref{gen\_inst}, \ref{headings}, and \ref{others} below.

% %% \subsection{Keywords for paper submission}

% %% Your NIPS paper can be submitted with any of the following keywords (more than one keyword is possible for each paper):

% %% \begin{verbatim}

% %% Bioinformatics

% %% Biological Vision

% %% Brain Imaging and Brain Computer Interfacing

% %% Clustering

% %% Cognitive Science

% %% Control and Reinforcement Learning

% %% Dimensionality Reduction and Manifolds

% %% Feature Selection

% %% Gaussian Processes

% %% Graphical Models

% %% Hardware Technologies

% %% Kernels

% %% Learning Theory

% %% Machine Vision

% %% Margins and Boosting

% %% Neural Networks

% %% Neuroscience

% %% Other Algorithms and Architectures

% %% Other Applications

% %% Semi-supervised Learning

% %% Speech and Signal Processing

% %% Text and Language Applications

% %% \end{verbatim}

% \section{General formatting instructions}

% \label{gen\_inst}

% The text must be confined within a rectangle 5.5~inches (33~picas) wide and

% 9~inches (54~picas) long. The left margin is 1.5~inch (9~picas).

% Use 10~point type with a vertical spacing of 11~points. Times New Roman is the

% preferred typeface throughout. Paragraphs are separated by 1/2~line space,

% with no indentation.

% Paper title is 17~point, initial caps/lower case, bold, centered between

% 2~horizontal rules. Top rule is 4~points thick and bottom rule is 1~point

% thick. Allow 1/4~inch space above and below title to rules. All pages should

% start at 1~inch (6~picas) from the top of the page.

% %The version of the paper submitted for review should have ``Anonymous Author(s)'' as the author of the paper.

% For the final version, authors' names are

% set in boldface, and each name is centered above the corresponding

% address. The lead author's name is to be listed first (left-most), and

% the co-authors' names (if different address) are set to follow. If

% there is only one co-author, list both author and co-author side by side.

% Please pay special attention to the instructions in section \ref{others}

% regarding figures, tables, acknowledgments, and references.

% \section{Headings: first level}

% \label{headings}

% First level headings are lower case (except for first word and proper nouns),

% flush left, bold and in point size 12. One line space before the first level

% heading and 1/2~line space after the first level heading.

% \subsection{Headings: second level}

% Second level headings are lower case (except for first word and proper nouns),

% flush left, bold and in point size 10. One line space before the second level

% heading and 1/2~line space after the second level heading.

% \subsubsection{Headings: third level}

% Third level headings are lower case (except for first word and proper nouns),

% flush left, bold and in point size 10. One line space before the third level

% heading and 1/2~line space after the third level heading.

% \section{Citations, figures, tables, references}

% \label{others}

% These instructions apply to everyone, regardless of the formatter being used.

% \subsection{Citations within the text}

% Citations within the text should be numbered consecutively. The corresponding

% number is to appear enclosed in square brackets, such as [1] or [2]-[5]. The

% corresponding references are to be listed in the same order at the end of the

% paper, in the \textbf{References} section. (Note: the standard

% \textsc{Bib\TeX} style \texttt{unsrt} produces this.) As to the format of the

% references themselves, any style is acceptable as long as it is used

% consistently.

% As submission is double blind, refer to your own published work in the

% third person. That is, use ``In the previous work of Jones et al.\ [4]'',

% not ``In our previous work [4]''. If you cite your other papers that

% are not widely available (e.g.\ a journal paper under review), use

% anonymous author names in the citation, e.g.\ an author of the

% form ``A.\ Anonymous''.

% \subsection{Footnotes}

% Indicate footnotes with a number\footnote{Sample of the first footnote} in the

% text. Place the footnotes at the bottom of the page on which they appear.

% Precede the footnote with a horizontal rule of 2~inches

% (12~picas).\footnote{Sample of the second footnote}

% \subsection{Figures}

% All artwork must be neat, clean, and legible. Lines should be dark

% enough for purposes of reproduction; art work should not be

% hand-drawn. The figure number and caption always appear after the

% figure. Place one line space before the figure caption, and one line

% space after the figure. The figure caption is lower case (except for

% first word and proper nouns); figures are numbered consecutively.

% Make sure the figure caption does not get separated from the figure.

% Leave sufficient space to avoid splitting the figure and figure caption.

% You may use color figures.

% However, it is best for the

% figure captions and the paper body to make sense if the paper is printed

% either in black/white or in color.

% \begin{figure}[h]

% \begin{center}

% %\framebox[4.0in]{$\;$}

% \fbox{\rule[-.5cm]{0cm}{4cm} \rule[-.5cm]{4cm}{0cm}}

% \end{center}

% \caption{Sample figure caption.}

% \end{figure}

% \subsection{Tables}

% All tables must be centered, neat, clean and legible. Do not use hand-drawn

% tables. The table number and title always appear before the table. See

% Table~\ref{sample-table}.

% Place one line space before the table title, one line space after the table

% title, and one line space after the table. The table title must be lower case

% (except for first word and proper nouns); tables are numbered consecutively.

% \begin{table}[t]

% \caption{Sample table title}

% \label{sample-table}

% \begin{center}

% \begin{tabular}{ll}

% \multicolumn{1}{c}{\bf PART} &\multicolumn{1}{c}{\bf DESCRIPTION}

% \\ \hline \\

% Dendrite &Input terminal \\

% Axon &Output terminal \\

% Soma &Cell body (contains cell nucleus) \\

% \end{tabular}

% \end{center}

% \end{table}

% \section{Final instructions}

% Do not change any aspects of the formatting parameters in the style files.

% In particular, do not modify the width or length of the rectangle the text

% should fit into, and do not change font sizes (except perhaps in the

% \textbf{References} section; see below). Please note that pages should be

% numbered.

% \section{Preparing PostScript or PDF files}

% Please prepare PostScript or PDF files with paper size ``US Letter'', and

% not, for example, ``A4''. The -t

% letter option on dvips will produce US Letter files.

% Fonts were the main cause of problems in the past years. Your PDF file must

% only contain Type 1 or Embedded TrueType fonts. Here are a few instructions

% to achieve this.

% \begin{itemize}

% \item You can check which fonts a PDF files uses. In Acrobat Reader,

% select the menu Files$>$Document Properties$>$Fonts and select Show All Fonts. You can

% also use the program \verb+pdffonts+ which comes with \verb+xpdf+ and is

% available out-of-the-box on most Linux machines.

% \item The IEEE has recommendations for generating PDF files whose fonts

% are also acceptable for NIPS. Please see

% \url{http://www.emfield.org/icuwb2010/downloads/IEEE-PDF-SpecV32.pdf}

% \item LaTeX users:

% \begin{itemize}

% \item Consider directly generating PDF files using \verb+pdflatex+

% (especially if you are a MiKTeX user).

% PDF figures must be substituted for EPS figures, however.

% \item Otherwise, please generate your PostScript and PDF files with the following commands:

% \begin{verbatim}

% dvips mypaper.dvi -t letter -Ppdf -G0 -o mypaper.ps

% ps2pdf mypaper.ps mypaper.pdf

% \end{verbatim}

% Check that the PDF files only contains Type 1 fonts.

% %For the final version, please send us both the Postscript file and

% %the PDF file.

% \item xfig "patterned" shapes are implemented with

% bitmap fonts. Use "solid" shapes instead.

% \item The \verb+\bbold+ package almost always uses bitmap

% fonts. You can try the equivalent AMS Fonts with command

% \begin{verbatim}

% \usepackage[psamsfonts]{amssymb}

% \end{verbatim}

% or use the following workaround for reals, natural and complex:

% \begin{verbatim}

% \newcommand{\RR}{I\!\!R} %real numbers

% \newcommand{\Nat}{I\!\!N} %natural numbers

% \newcommand{\CC}{I\!\!\!\!C} %complex numbers

% \end{verbatim}

% \item Sometimes the problematic fonts are used in figures

% included in LaTeX files. The ghostscript program \verb+eps2eps+ is the simplest

% way to clean such figures. For black and white figures, slightly better

% results can be achieved with program \verb+potrace+.

% \end{itemize}

% \item MSWord and Windows users (via PDF file):

% \begin{itemize}

% \item Install the Microsoft Save as PDF Office 2007 Add-in from

% \url{http://www.microsoft.com/downloads/details.aspx?displaylang=en\&familyid=4d951911-3e7e-4ae6-b059-a2e79ed87041}

% \item Select ``Save or Publish to PDF'' from the Office or File menu

% \end{itemize}

% \item MSWord and Mac OS X users (via PDF file):

% \begin{itemize}

% \item From the print menu, click the PDF drop-down box, and select ``Save

% as PDF...''

% \end{itemize}

% \item MSWord and Windows users (via PS file):

% \begin{itemize}

% \item To create a new printer

% on your computer, install the AdobePS printer driver and the Adobe Distiller PPD file from

% \url{http://www.adobe.com/support/downloads/detail.jsp?ftpID=204} {\it Note:} You must reboot your PC after installing the

% AdobePS driver for it to take effect.

% \item To produce the ps file, select ``Print'' from the MS app, choose

% the installed AdobePS printer, click on ``Properties'', click on ``Advanced.''

% \item Set ``TrueType Font'' to be ``Download as Softfont''

% \item Open the ``PostScript Options'' folder

% \item Select ``PostScript Output Option'' to be ``Optimize for Portability''

% \item Select ``TrueType Font Download Option'' to be ``Outline''

% \item Select ``Send PostScript Error Handler'' to be ``No''

% \item Click ``OK'' three times, print your file.

% \item Now, use Adobe Acrobat Distiller or ps2pdf to create a PDF file from

% the PS file. In Acrobat, check the option ``Embed all fonts'' if

% applicable.

% \end{itemize}

% \end{itemize}

% If your file contains Type 3 fonts or non embedded TrueType fonts, we will

% ask you to fix it.

% \subsection{Margins in LaTeX}

% Most of the margin problems come from figures positioned by hand using

% \verb+\special+ or other commands. We suggest using the command

% \verb+\includegraphics+

% from the graphicx package. Always specify the figure width as a multiple of

% the line width as in the example below using .eps graphics

% \begin{verbatim}

% \usepackage[dvips]{graphicx} ...

% \includegraphics[width=0.8\linewidth]{myfile.eps}

% \end{verbatim}

% or % Apr 2009 addition

% \begin{verbatim}

% \usepackage[pdftex]{graphicx} ...

% \includegraphics[width=0.8\linewidth]{myfile.pdf}

% \end{verbatim}

% for .pdf graphics.

% See section 4.4 in the graphics bundle documentation (\url{http://www.ctan.org/tex-archive/macros/latex/required/graphics/grfguide.ps})

% A number of width problems arise when LaTeX cannot properly hyphenate a

% line. Please give LaTeX hyphenation hints using the \verb+\-+ command.

% \subsubsection\*{Acknowledgments}

% Use unnumbered third level headings for the acknowledgments. All

% acknowledgments go at the end of the paper. Do not include

% acknowledgments in the anonymized submission, only in the

% final paper.

% \subsubsection\*{References}

% References follow the acknowledgments. Use unnumbered third level heading for

% the references. Any choice of citation style is acceptable as long as you are

% consistent. It is permissible to reduce the font size to `small' (9-point)

% when listing the references. {\bf Remember that this year you can use

% a ninth page as long as it contains \emph{only} cited references.}

% \small{

% [1] Alexander, J.A. \& Mozer, M.C. (1995) Template-based algorithms

% for connectionist rule extraction. In G. Tesauro, D. S. Touretzky

% and T.K. Leen (eds.), {\it Advances in Neural Information Processing

% Systems 7}, pp. 609-616. Cambridge, MA: MIT Press.

% [2] Bower, J.M. \& Beeman, D. (1995) {\it The Book of GENESIS: Exploring

% Realistic Neural Models with the GEneral NEural SImulation System.}

% New York: TELOS/Springer-Verlag.

% [3] Hasselmo, M.E., Schnell, E. \& Barkai, E. (1995) Dynamics of learning

% and recall at excitatory recurrent synapses and cholinergic modulation

% in rat hippocampal region CA3. {\it Journal of Neuroscience}

% {\bf 15}(7):5249-5262.

% }

\end{document}