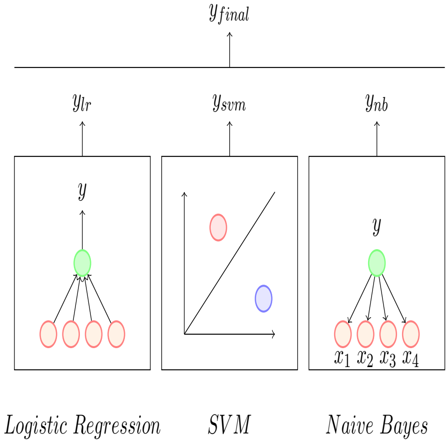
Module 8.10 : Ensemble methods

Other forms of regularization

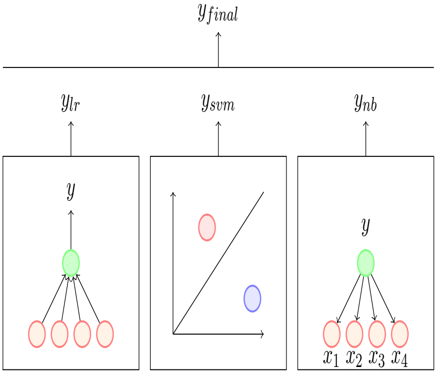
- l_2 regularization
- Dataset augmentation
- Parameter Sharing and tying
- Adding Noise to the inputs
- Adding Noise to the outputs
- Early stopping
- Ensemble methods
- Dropout

Other forms of regularization

- l_2 regularization
- Dataset augmentation
- Parameter Sharing and tying
- Adding Noise to the inputs
- Adding Noise to the outputs
- Early stopping
- Ensemble methods
- Dropout

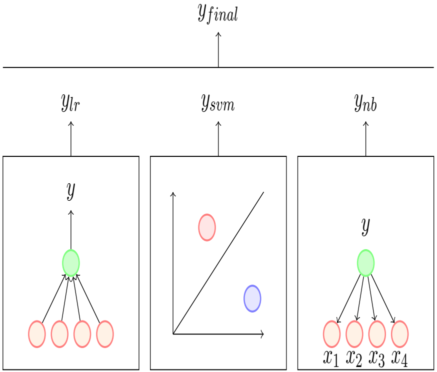


- Combine the output of different models to reduce generalization error



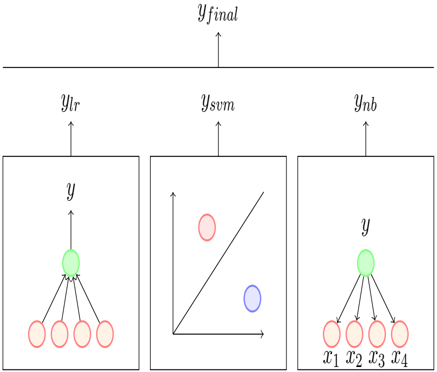
Logistic Regression *SVM* *Naive Bayes*

- Combine the output of different models to reduce generalization error
- The models can correspond to different classifiers



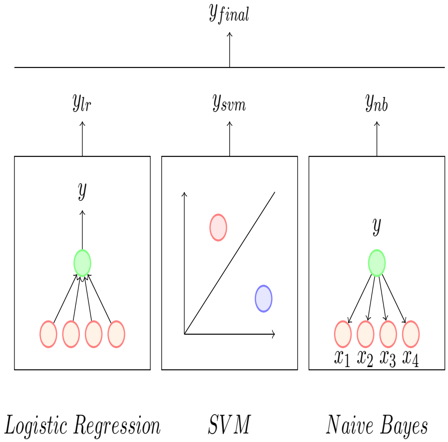
Logistic Regression *SVM* *Naive Bayes*

- Combine the output of different models to reduce generalization error
- The models can correspond to different classifiers
- It could be different instances of the same classifier trained with:

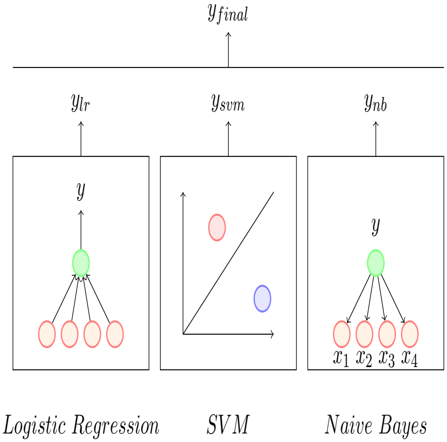


Logistic Regression *SVM* *Naive Bayes*

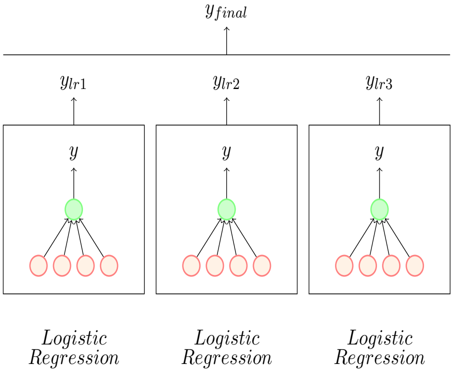
- Combine the output of different models to reduce generalization error
- The models can correspond to different classifiers
- It could be different instances of the same classifier trained with:
  - different hyperparameters

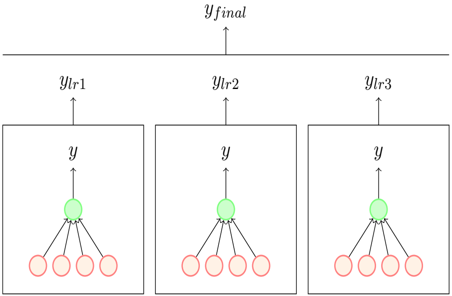


- Combine the output of different models to reduce generalization error
- The models can correspond to different classifiers
- It could be different instances of the same classifier trained with:
  - different hyperparameters
  - different features

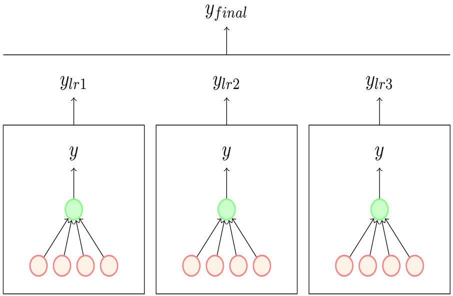


- Combine the output of different models to reduce generalization error
- The models can correspond to different classifiers
- It could be different instances of the same classifier trained with:
  - different hyperparameters
  - different features
  - different samples of the training data


y_{final}
y_{lr1}
y_{lr2}
y_{lr3}
y
y
y
Logistic Regression
Logistic Regression
Logistic Regression

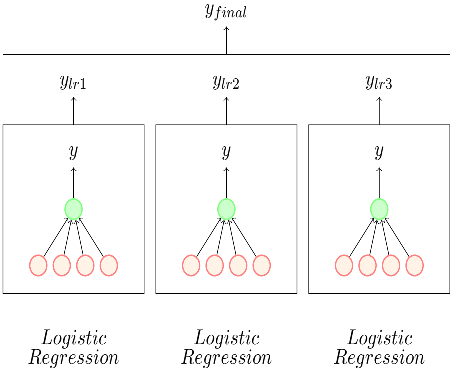


Logistic Regression *Logistic Regression* *Logistic Regression*

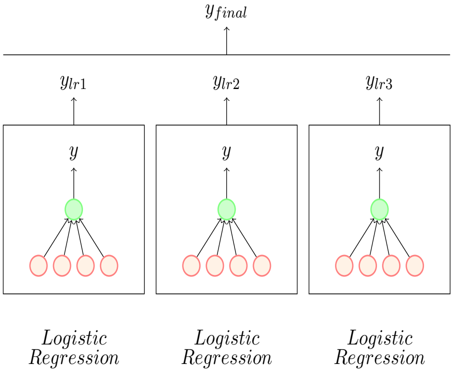


Logistic Regression *Logistic Regression* *Logistic Regression*

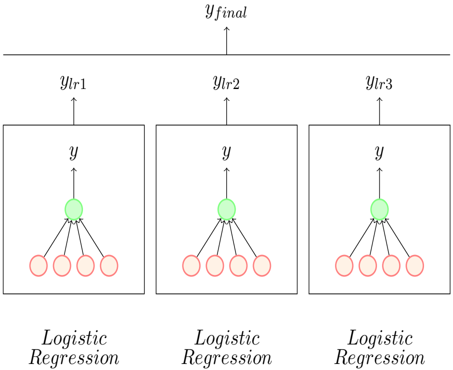
- Bagging: form an ensemble using different instances of the same classifier



- Bagging: form an ensemble using different instances of the same classifier
- From a given dataset, construct multiple training sets by sampling with replacement $(T_1, T_2, ..., T_k)$



- Bagging: form an ensemble using different instances of the same classifier
- From a given dataset, construct multiple training sets by sampling with replacement $(T_1, T_2, ..., T_k)$
- Train i^{th} instance of the classifier using training set T_i



Each model trained with a different sample of the data (sampling with replacement)

- Bagging: form an ensemble using different instances of the same classifier
- From a given dataset, construct multiple training sets by sampling with replacement $(T_1, T_2, ..., T_k)$
- Train i^{th} instance of the classifier using training set T_i

- When would bagging work?

- When would bagging work?
- Consider a set of k LR models

- When would bagging work?
- Consider a set of k LR models
- Suppose that each model makes an error ε_i on a test example

- When would bagging work?
- Consider a set of k LR models
- Suppose that each model makes an error ε_i on a test example
- Let ε_i be drawn from a zero mean multivariate normal distribution

- When would bagging work?
- Consider a set of k LR models
- Suppose that each model makes an error ε_i on a test example
- Let ε_i be drawn from a zero mean multivariate normal distribution
- $Variance = E[\varepsilon_i^2] = V$

- When would bagging work?
- Consider a set of k LR models
- Suppose that each model makes an error ε_i on a test example
- Let ε_i be drawn from a zero mean multivariate normal distribution
- $Variance = E[\varepsilon_i^2] = V$
- $Covariance = E[\varepsilon_i \varepsilon_j] = C$

- The error made by the average prediction of all the models is $\frac{1}{k}\sum_i \varepsilon_i$

- When would bagging work?
- Consider a set of k LR models
- Suppose that each model makes an error ε_i on a test example
- Let ε_i be drawn from a zero mean multivariate normal distribution
- $Variance = E[\varepsilon_i^2] = V$
- $Covariance = E[\varepsilon_i \varepsilon_j] = C$

- The error made by the average prediction of all the models is $\frac{1}{k}\sum_i \varepsilon_i$
- The expected squared error is :

- When would bagging work?
- Consider a set of k LR models
- Suppose that each model makes an error ε_i on a test example
- Let ε_i be drawn from a zero mean multivariate normal distribution
- $Variance = E[\varepsilon_i^2] = V$
- $Covariance = E[\varepsilon_i \varepsilon_j] = C$

- The error made by the average prediction of all the models is $\frac{1}{k}\sum_i \varepsilon_i$
- The expected squared error is :

$$mse = E[(\frac{1}{k}\sum_i \varepsilon_i)^2]$$

- When would bagging work?
- Consider a set of k LR models
- Suppose that each model makes an error ε_i on a test example
- Let ε_i be drawn from a zero mean multivariate normal distribution
- $Variance = E[\varepsilon_i^2] = V$
- $Covariance = E[\varepsilon_i \varepsilon_j] = C$

- The error made by the average prediction of all the models is $\frac{1}{k}\sum_i \varepsilon_i$
- The expected squared error is :

$$
\begin{aligned}
mse =& E[(\frac{1}{k}\sum_i \varepsilon_i)^2] \\
=& \frac{1}{k^2} E[\sum_i \sum_{i=j} \varepsilon_i \varepsilon_j + \sum_i \sum_{i \neq j} \varepsilon_i \varepsilon_j]
\end{aligned}
$$

- When would bagging work?
- Consider a set of k LR models
- Suppose that each model makes an error ε_i on a test example
- Let ε_i be drawn from a zero mean multivariate normal distribution
- $Variance = E[\varepsilon_i^2] = V$
- $Covariance = E[\varepsilon_i \varepsilon_j] = C$

- The error made by the average prediction of all the models is $\frac{1}{k}\sum_i \varepsilon_i$
- The expected squared error is :

$$
\begin{aligned}
mse =& E[(\frac{1}{k}\sum_i \varepsilon_i)^2] \\
=& \frac{1}{k^2} E[\sum_i \sum_{i=j} \varepsilon_i \varepsilon_j + \sum_i \sum_{i \neq j} \varepsilon_i \varepsilon_j] \\
=& \frac{1}{k^2} E[\sum_i \varepsilon_i^2 + \sum_i \sum_{i \neq j} \varepsilon_i \varepsilon_j]
\end{aligned}
$$

- When would bagging work?
- Consider a set of k LR models
- Suppose that each model makes an error ε_i on a test example
- Let ε_i be drawn from a zero mean multivariate normal distribution
- $Variance = E[\varepsilon_i^2] = V$
- $Covariance = E[\varepsilon_i \varepsilon_j] = C$

- The error made by the average prediction of all the models is $\frac{1}{k}\sum_i \varepsilon_i$
- The expected squared error is :

$$
\begin{aligned}
mse =& E[(\frac{1}{k}\sum_i \varepsilon_i)^2] \\
=& \frac{1}{k^2}E[\sum_i \sum_{i=j} \varepsilon_i \varepsilon_j + \sum_i \sum_{i \neq j} \varepsilon_i \varepsilon_j] \\
=& \frac{1}{k^2}E[\sum_i \varepsilon_i^2 + \sum_i \sum_{i \neq j} \varepsilon_i \varepsilon_j] \\
=& \frac{1}{k^2}(\sum_i E[\varepsilon_i^2] + \sum_i \sum_{i \neq j} E[\varepsilon_i \varepsilon_j])
\end{aligned}
$$

- When would bagging work?
- Consider a set of k LR models
- Suppose that each model makes an error ε_i on a test example
- Let ε_i be drawn from a zero mean multivariate normal distribution
- $Variance = E[\varepsilon_i^2] = V$
- $Covariance = E[\varepsilon_i \varepsilon_j] = C$

- When would bagging work?
- Consider a set of k LR models
- Suppose that each model makes an error ε_i on a test example
- Let ε_i be drawn from a zero mean multivariate normal distribution
- $Variance = E[\varepsilon_i^2] = V$
- $Covariance = E[\varepsilon_i \varepsilon_j] = C$
- The error made by the average prediction of all the models is $\frac{1}{k}\sum_i \varepsilon_i$
- The expected squared error is :

$$
\begin{aligned}
mse =& E[(\frac{1}{k}\sum_i \varepsilon_i)^2] \\
=& \frac{1}{k^2} E[\sum_i \sum_{i=j} \varepsilon_i \varepsilon_j + \sum_i \sum_{i \neq j} \varepsilon_i \varepsilon_j] \\
=& \frac{1}{k^2} E[\sum_i \varepsilon_i^2 + \sum_i \sum_{i \neq j} \varepsilon_i \varepsilon_j] \\
=& \frac{1}{k^2} (\sum_i E[\varepsilon_i^2] + \sum_i \sum_{i \neq j} E[\varepsilon_i \varepsilon_j]) \\
=& \frac{1}{k^2} (kV + k(k-1)C)
\end{aligned}
$$

- The error made by the average prediction of all the models is $\frac{1}{k}\sum_i \varepsilon_i$
- The expected squared error is :

$$
\begin{aligned}
mse &= E[(\frac{1}{k}\sum_i \varepsilon_i)^2] \\
&= \frac{1}{k^2}E[\sum_i \sum_{i=j} \varepsilon_i\varepsilon_j + \sum_i \sum_{i\neq j} \varepsilon_i\varepsilon_j] \\
&= \frac{1}{k^2}E[\sum_i \varepsilon_i^2 + \sum_i \sum_{i\neq j} \varepsilon_i\varepsilon_j] \\
&= \frac{1}{k^2}(\sum_i E[\varepsilon_i^2] + \sum_i \sum_{i\neq j} E[\varepsilon_i\varepsilon_j]) \\
&= \frac{1}{k^2}(kV + k(k-1)C) \\
&= \frac{1}{k}V + \frac{k-1}{k}C
\end{aligned}
$$

- When would bagging work?
- Consider a set of k LR models
- Suppose that each model makes an error ε_i on a test example
- Let ε_i be drawn from a zero mean multivariate normal distribution
- $Variance = E[\varepsilon_i^2] = V$
- $Covariance = E[\varepsilon_i\varepsilon_j] = C$

$$mse = \frac{1}{k}V + \frac{k-1}{k}C$$

$$mse = \frac{1}{k}V + \frac{k-1}{k}C$$

- When would bagging work ?

$$mse = \frac{1}{k}V + \frac{k-1}{k}C$$

- When would bagging work ?
- If the errors of the model are perfectly correlated then $V = C$ and $mse = V$ [bagging does not help: the mse of the ensemble is as bad as the individual models]

$$mse = \frac{1}{k}V + \frac{k-1}{k}C$$

- When would bagging work ?
- If the errors of the model are perfectly correlated then $V = C$ and $mse = V$ [bagging does not help: the mse of the ensemble is as bad as the individual models]
- If the errors of the model are independent or uncorrelated then $C = 0$ and the mse of the ensemble reduces to $\frac{1}{k}V$

$$mse = \frac{1}{k}V + \frac{k-1}{k}C$$

- When would bagging work ?
- If the errors of the model are perfectly correlated then $V = C$ and $mse = V$ [bagging does not help: the mse of the ensemble is as bad as the individual models]
- If the errors of the model are independent or uncorrelated then $C = 0$ and the mse of the ensemble reduces to $\frac{1}{k}V$
- On average, the ensemble will perform at least as well as its individual members