

figure 3

Після чого описується процес навчання. Зазначається, що воно відбувається на одному графічному процесорі. Далі у процесі навчання послідовно видається таблиця з характеристиками процесу навчання: Epoch – номер виконаної епохи, Iteration – номер поточної ітерації, Time Elapsed – обсяг витраченого часу, Mini-batch loss – втрати для мініпакету (підвибірки), Mini-batch Accuracy – точність для мініпакету, Base Learning Rate – базовий крок навчання (коригувальний приріст).

Далі візуалізуються тестові послідовності (figure 3). Після чого видаються опис трьох елементів тестової вибірки та точність accuracy = 0.9432.

## 6.2 Моделювання нейронних мереж у пакеті Statistica Neural Networks

*Statistica Neural Networks* – це універсальний пакет нейромережевого аналізу фірми StatSoft. Він може працювати як самостійний додаток, а також у рамках системи *Statistica*. Пакет має ряд переваг.

Він містить найсучасніші і могутні методи навчання мережі (включаючи методи сполучених градієнтів та Левенберга-Марквардта); має можливість створювати складні, практично не обмежені в розмірах комбінації з мереж різних архітектур; дозволяє вибіркоче навчання фрагментів нейронної мережі; забезпечує повний контроль над вибором функцій активації; містить систему *Network Wizard* – Нейро-Майстер, що допомагає користувачу приймати рішення і вибирати параметри.

Пакет має виняткову простоту у використанні та неперевершену аналітичну міць. Так, наприклад, *Automatic Network Designer* – Автоматичний Конструктор Мережі знайде найкращу архітектуру для НМ (вибір потрібної архітектури мережі – це важкий процес проб і помилок; Statistica Neural Networks зробить це за користувача).

Пакет містить могутні методи аналізу, у тому числі, унікальний інструмент нейро-генетичного відбору входних даних – *Neuro-Genetic Input Selection* (вибір потрібних входних змінних при розвідницькому аналізі даних – області, де найчастіше застосовуються нейромережі – займає багато часу; пакет Statistica Neural Networks здатний виконувати цю роботу за користувача).

Statistica Neural Networks може працювати у середовищі Statistica, але також може використовуватися і як самостійний додаток (є можливість уведення-виведення файлів даних системи Statistica; матриць вихідних даних і графіки).

Простий і зручний інтерфейс системи Statistica Neural Networks дозволяє швидко створювати нейромережеві додатки для вирішення користувальницьких задач.

Statistica Neural Networks цілком підтримує *інтерфейс прикладного програмування* – API, так що досвідчені користувачі (і розроблювачі користувальницьких систем «пошуку знань» і «видобутку даних») можуть використовувати могутні обчислювальні можливості модуля Statistica Neural Networks у своїх власних додатках. Інтерфейс прикладного програмування Statistica Neural Networks' API дозволяє вмонтувати ці додатки у вже наявну в систему, наприклад, зробити їх частиною більш широкого обчислювального середовища (це можуть бути процедури, розроблені окремо й

убудовані в комп'ютерну мережу фірми). API поставляється в двох варіантах. Скорочена версія API дозволяє запускати створені і навчені в середовищі Statistica Neural Networks нейронні мережі з інших програм-додатків, написаних на С, С++, Delphi або Visual Basic. Повний API надає користувачу доступ до всіх можливостей ядра пакета Statistica Neural Networks, включаючи створення, редагування, введення/виведення даних і архітектур мереж, навчання мереж.

Пакет має багатовіконний Windows інтерфейс. Головне і дочірнє вікна пакета зображені на рис. 6.19.

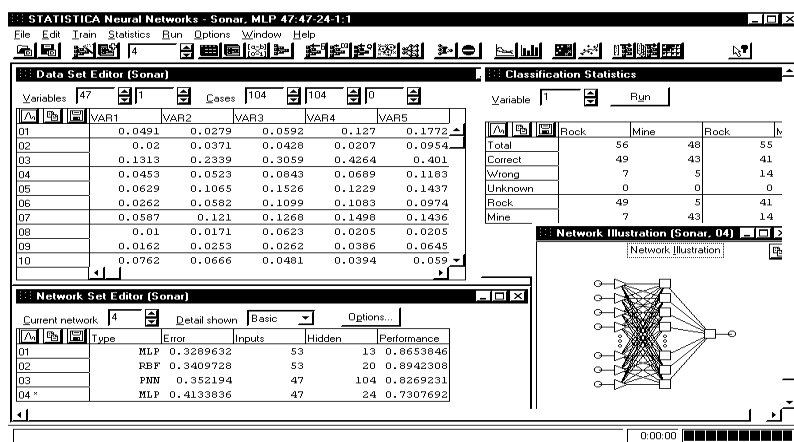


Рисунок 6.19 – Головне вікно пакета Statistica Neural Networks

Розглянемо призначення основних пунктів меню пакета.

Меню File (Файл) містить пункти: New – новий(ва) (Network – нейронна мережа, Intelligent Problem Solver – інтелектуальний вирішувач задач, Data Set – вибірка даних); Open – відкрити файл, Close – закрити файл, Save – зберегти в поточний файл, Save as – зберегти у файл з ім'ям; Network Set – набір нейромереж; Exit – вихід.

Меню Edit (Редагувати) містить пункти: Variables – змінні, ознаки; Cases – випадки, екземпляри; Undo – скасувати дію; Cut – вирізати; Copy – копіювати до буфера; Paste – вставити з буфера; Clear – очистити; Data Set – набір даних; Network Set – набір

нейромереж; Pre/post Processing – попередня / пост-обробка; Network – нейромережа.

Меню Train (Навчання) містить пункти: Multilayer Perceptrons – багатошарові перцептрони; Radial-Basis Functions – радіально-базисні мережі; Linear – лінійні мережі; Kohonen – мережі Кохонена (SOM); Class labels – мітки класів; LVQ – мережі LVQ; Probabilistic – імовірнісні мережі; General Regression – узагальнено-регресійні мережі; Principal Components – аналіз головних компонент; Auxiliary – помічник; Best Network – добір кращої мережі, Stopping Conditions – критерії зупинення, Set Weights – установка wag, Weigend Regularization – регуляція wag, Feature Selection – відбір ознак.

Меню Statistics (Статистика) містить пункти: Training Graph – графік навчання; Case Errors – помилки для екземплярів; Classification – класифікація; ROC Curve – крива характеристики дії одержувача; Regression – регресія; Weight Distribution – розподіл wag; Win Frequencies – частоти переможців; Sensivity – чутливість; Network Illustration – зображення мережі; Messages – повідомлення.

Меню Run (Запуск) містить пункти: Single Case – запуск для одного екземпляра; Data Set – запуск для набору даних; One-off – запуск для одного з прикладів, обраних інтерактивно; Activations – активація мережі; Time Series – часові ряди; Response Graph – графік відгуку; Response Surface – поверхня відгуку; Topological Map – топологічна карта; Cluster Diagram – діаграма кластерів; Code Generation – генерація коду.

Вибір вхідних змінних – це винятково важливий етап при побудові нейронної мережі. Перед тим, як безпосередньо починати працювати з пакетом Statistica Neural Networks, має сенс зробити попередній відбір змінних, використовуючи при цьому свої знання в предметній області та стандартні статистичні критерії. Потім, уже засобами пакета Statistica Neural Networks можна буде спробувати різні комбінації вхідних змінних.

У пакеті Statistica Neural Networks є можливість «ігнорувати» деякі змінні, так що отримана мережа не буде використовувати їх як входи. Можна по черзі

експериментувати з різними комбінаціями входів, будуючи кожного разу нові варіанти мереж.

*Процес побудови нейромережі* (після вибору вхідних змінних) складається з таких кроків.

1. Вибрати початкову конфігурацію мережі (наприклад, один проміжний шар з числом елементів у ньому, рівним півсумі кількості входів і кількості виходів – Наставник (Network Advisor) пакета Statistica Neural Networks запропонує таку конфігурацію за замовчуванням).

2. Провести ряд експериментів з різними конфігураціями, запам'ятовуючи при цьому кращу мережу (у змісті контрольної помилки). У пакеті Statistica Neural Networks передбачене автоматичне запам'ятовування кращої мережі під час експерименту. Для кожної конфігурації варто провести кілька експериментів, щоб не одержати помилковий результат через те, що метод навчання потрапив у локальний мінімум цільової функції.

3. Якщо в черговому експерименті спостерігається недонавчання (мережа не видає результат прийнятної якості), спробувати додати додаткові нейрони в проміжний шар (шари). Якщо це не допомагає, спробувати додати новий проміжний шар.

4. Якщо має місце перенавчання (контрольна помилка стала зростати), спробувати видалити кілька схованих елементів (а можливо і шарів).

Багаторазове повторення евристичних експериментів у кращому випадку є досить утомливим, і тому в пакет Statistica Neural Networks включений спеціальний метод автоматичного пошуку.

Автоматичний конструктор мережі – *Automatic Network Designer* проведе експерименти з різним числом схованих елементів, для кожної спробної архітектури мережі виконає кілька прогонів навчання, відбираючи при цьому найкращу мережу по показнику контрольної помилки з виправленням на розмір мережі. В Автоматичному конструкторі мережі реалізовані складні методи пошуку, у тому числі метод «імітації відпалу» (simulated annealing), за допомогою яких можна

перепробувати сотні різних мереж, виділяючи з них особливо перспективні, або швидко знаходити «грубе і просте» рішення.

У задачі класифікації мережа повинна віднести кожне спостереження до одного з декількох класів (чи, у більш загальному випадку, оцінити імовірність приналежності спостереження до кожного з класів). У пакеті Statistica Neural Networks для класифікації використовуються номінальна вихідна змінна – різні її значення відповідають різним класам.

У пакеті Statistica Neural Networks класифікацію можна здійснювати за допомогою мереж наступних типів: багатосарового перцептрона, радіально-базисних функцій, мережі Кохонена, імовірнісної нейромережі та лінійної мережі. Єдина з мереж пакета Statistica Neural Networks, не призначена для задач класифікації, – це узагальнено-регресійна мережа.

Номінальні змінні представляються в пакеті Statistica Neural Networks в одному з двох видів (перший з них підходить тільки для змінних із двома значеннями): 1) бінарному (два стани) та 2) один-з-N. При бінарному поданні змінній відповідає один вузол мережі, при цьому значення 0,0 означає активний стан, а 1,0 – неактивний. При кодуванні 1-з-N на кожен стан виділяється один елемент, так що кожен конкретний стан представляється як 1,0 у відповідному елементі та 0,0 у всіх інших.

Номінальні вхідні змінні в пакеті Statistica Neural Networks можуть бути перетворені одним з цих методів як на етапі навчання мережі, так і при її роботі. Цільові вихідні значення для елементів, що відповідають номінальним змінним, також легко визначаються під час навчання. Більш великі зусилля вимагаються для того, щоб за результатами роботи мережі визначити вихідний клас.

Кожний з вихідних елементів буде містити числові значення в інтервалі від 0,0 до 1,0. Щоб упевнено визначити клас по наборі вихідних значень, мережа повинна вирішити, чи «достить близькі» вони до нуля або одиниці. Якщо такої близькості не спостерігається, клас вважається «невизначеним».

Крім того, у пакеті Statistica Neural Networks для інтерпретації вихідних значень використовуються довірчі рівні (пороги прийняття і відкидання). Ці граничні значення можна корегувати, щоб змусити мережу бути більш, чи, навпаки, менш

«рішучою» при оголошенні класу. Схеми тут несильно розрізняються для випадків бінарного та 1-з-N подань.

Бінарне подання. Якщо вихідне значення елемента перевищує поріг прийняття, то вибирається клас 1,0. Якщо вихідне значення лежить нижче порога відкидання, вибирається клас 0,0. Якщо вихідне значення лежить між порогами, клас вважається не визначеним.

Один-з-N. Визначений клас вибирається тільки в тому випадку, якщо значення відповідного вихідного елемента вище порога прийняття, а всіх інших вихідних елементів – нижче порога відкидання. Якщо ж дана умова не виконана, клас не визначається.

При кодуванні методом 1-з-N є одна особливість. На перший погляд здається, що «найбільш рішучою» буде мережа з порогами прийняття і відкидання, рівними 0,5. Це дійсно так для бінарного кодування, але вже не так для кодування 1-з-N. Можна зробити так, щоб поріг прийняття був нижче порога відкидання, і найбільш рішучою буде мережа, у якій поріг прийняття 0,0, а поріг відкидання 1,0. При такому, на перший погляд, дивному настроюванню мережі елемент із найвищим рівнем активації буде визначати клас поза залежністю від того, що відбувається в інших елементах.

Схема дії *методу визначення класу* в пакеті Statistica Neural Networks полягає у виконанні таких кроків.

1. Вибирається елемент із найвищим вихідним сигналом. Якщо його вихідний сигнал вище або дорівнює порогові прийняття, а вихідні сигнали всіх інших елементів нижче порога відкидання, то як відповідь видати клас, обумовлений цим елементом.

2. При порозі прийняття 0,0 вихідний сигнал елемента, що виграв, завжди буде прийнятий, а при порозі відкидання 1,0 всі інші елементи неминуче будуть відкинуті, і тому метод зводиться до простого вибору елемента, що виграв. Якщо ж обое граничних значення – прийняття та відкидання – установити на рівні 0,5, мережа цілком може залишитися в нерішучості (у випадках, коли в переможця результат нижче 0.5 чи в когось із тих, що програли – вище 0,5).

Усе сказане відноситься до механізму вибору класу для більшості типів мереж. На відміну від них, мережа Кохонена діє зовсім інакше.

У мережі Кохонена елементом топологічної карти (вихідного шару), що виграв, є той, у якого найвищий рівень активації (він вимірює відстань від вхідного приклада до точки, координати якої зберігаються в елементі мережі). Деякі чи навіть усі елементи топологічної карти можуть бути позначені іменами класів. Якщо ця відстань є достатньо малою, то даний випадок зараховується до відповідного класу (за умови, що зазначене ім'я класу). У пакеті Statistica Neural Networks значення порога прийняття – це найбільша відстань, на якій приймається позитивне рішення про класифікацію спостереження. Якщо ж вхідний випадок лежить від елемента, що виграв, на більш далекій відстані або якщо елемент, що виграв, не був позначений (або якщо його мітка не відповідає жодному зі значень вихідний номінальної змінної), то випадок залишається неklasифікованим. Поріг відкидання в мережах Кохонена не використовується.

У наших розглядах ми припускали, що «позитивному» рішення про класифікацію повинне відповідати значення, близьке до 1,0, а «негативному» – близьке до 0,0. Це дійсно так, якщо на виході використовуються логістичні функції активації. Крім того, це зручно, оскільки імовірність може приймати значення від 0,0 до 1,0. Однак, у деяких ситуаціях може виявитися більш зручним використовувати інший діапазон. Іноді застосовується зворотна упорядкованість, так що позитивне рішення відповідає малим вихідним значенням. Пакет Statistica Neural Networks підтримує кожної з цих варіантів роботи.

Спочатку як границі діапазону для кожної змінної використовуються значення мінімум/середнє і максимум/стандартне відхилення. Для логістичної вихідної функції активації гарними значеннями за замовчуванням є 0,0 і 1,0. Деякі автори радять використовувати як функцію активації гіперболічний тангенс, що приймає значення в інтервалі  $(-1,0; +1,0)$ . Таким прийомом можна поліпшити навчання, тому що ця функція (на відміну від логістичної) симетрична. У цьому випадку потрібно змінити значення мінімум/середнє і



максимум/стандартне відхилення, і програма Statistica Neural Networks автоматично буде правильно інтерпретувати класи.

Зворотна упорядкованість, як правило, застосовується в двох ситуаціях. Одна з них – це мережі Кохонена, у яких вихідне значення є міра далекості, і її мале значення відповідає більшій довірі. Друга ситуація виникає при використанні матриці втрат (яка може бути додана у імовірнісну мережу на етапі її побудови або вручну – до мереж інших типів). Якщо використовується матриця втрат, то виходи мережі означають очікувані утрати від вибору того чи іншого класу, і ціль полягає в тім, щоб вибрати клас з найменшими втратами. Упорядкованість можна звернути, оголосивши вихідний сигнал не рівнем довіри, а мірою помилки. У такому випадку поріг прийняття буде нижче порога відкидання.

При виборі порогів прийняття/відкидання й оцінці здібностей мережі до класифікації дуже допомагає інформація, що міститься у вікні Статистики класифікації – *Classification Statistics*. У ньому вказується, скільки спостережень було класифіковано правильно, скільки неправильно взагалі чи не класифіковано. Крім того, видається інформація про те, скільки спостережень кожного класу було віднесено до інших класів. Усі ці дані видаються окремо для навчальної, контрольної і тестової множин.

У задачах оцінювання (регресії) метою є оцінка значення числової вихідної змінної за значеннями вхідних змінних. Задачі регресії в пакеті Statistica Neural Networks можна вирішувати за допомогою: багат шарових перцептронів, мереж радіально-базисних функцій, узагальнено-регресійних і лінійних мереж. При цьому вихідні дані повинні мати стандартний числовий (не номінальний) тип.

Особливу важливість для регресії мають масштабування (шкалування) вхідних значень і ефекти екстраполяції.

Нейронні мережі найбільше часто використовуваних архітектур видають вихідні значення в деякому визначеному діапазоні (наприклад, на відрізку  $[0,1]$  у випадку логістичної функції активації). Для задач класифікації це не створює

труднощів. Однак для задач регресії зовсім очевидно, що отут є проблема, і деякі її деталі виявляються дуже тонкими.

Для початку застосуємо метод *масштабування*, щоб вихід мережі мав «прийнятний» діапазон. Найпростішою з масштабуючих функцій пакета Statistica Neural Networks є мінімаксна функція: вона знаходить мінімальне і максимальне значення змінної по навчальній множині і виконує лінійне перетворення (із застосуванням коефіцієнта масштабу і зсуву), так щоб значення лежали в потрібному діапазоні (як правило, на відрізьку

$[0,0; 1,0]$ ). Якщо ці дії застосовуються до числової вихідної змінної, тобто є гарантія, що всі навчальні значення після перетворення потраплять в область можливих вихідних значень мережі, тоді мережа може бути навчена. Крім того, ми знаємо, що виходи мережі повинні знаходитися у визначених межах. Цю обставина можна вважати достоїнством або недоліком – тут ми приходимо до питань екстраполяції.

Припустимо, наприклад, що ми використовуємо багат шаровий персептрон. Застосування мінімакса по описаній вище схемі є дуже обмеженим. По-перше, функція не буде екстраполюватися, як би близько ми не знаходилися до навчальних даних (у дійсності ж, якщо ми лише мало-мало вийшли за область навчальних даних, екстраполяція цілком виправдана). По-друге, оцінка по середньому також не буде виконуватися: замість цього буде братися мінімум або максимум дивлячись по тому, зростала чи убувала в цьому місці оцінювана крива.

Щоб уникнути цих недоліків у персептроні використовується ряд прийомів.

По-перше, логістичну функцію активації у вихідному шарі можна замінити на лінійну, котра не змінює рівня активації (помітимо, що функції активації змінюються тільки у вихідному шарі; у проміжних шарах як і раніше залишаються логістичні та гіперболічні функції активації). Лінійна функція активації не насичується, і тому здатна екстраполювати (при цьому логістичні функції попередніх рівнів усе-таки припускають насичення на більш високих рівнях). Лінійні функції активації у персептроні можуть викликати визначені обчислювальні труднощі в методі

зворотного поширення, тому при його використанні варто брати малі (менш 0,1) швидкості навчання. Описаний підхід придатний для цілей екстраполяції.

По-друге, можна змінити цільовий діапазон мінімаксної масштабуючої функції (наприклад, зробити його  $[0,25; 0,75]$ ). У результаті навчальні спостереження будуть відображатися в рівні, що відповідає середній частини діапазону вихідних значень. Цікаво помітити, що якщо цей діапазон обраний маленьким, і обидві його границі знаходяться поблизу значення 0,5, то він буде відповідати середній ділянці сигмоїдної кривої, на якому вона «майже лінійна», – тоді ми будемо мати практично ту ж схему, що й у випадку лінійного вихідного шару. Така мережа зможе виконувати екстраполяцію у визначених межах, а потім буде насичуватися. Усе це можна добре собі представити так: екстраполяція припустима у визначених границях, а поза ними вона буде припинятися.

Якщо застосовується перший підхід і у вихідному шарі поміщені лінійні елементи, то може вийти так, що взагалі немає необхідності використовувати метод масштабування, оскільки елементи і без масштабування можуть видавати будь-який рівень вихідних сигналів. У пакеті Statistica Neural Networks мається можливість для більшої ефективності узагалі відключити всі масштабування. Однак, на практиці повне відмовлення від масштабування приводить до труднощів в методах навчання. Дійсно, у цьому випадку різні ваги мережі працюють у різних масштабах, і це ускладнює початкову ініціалізацію ваг і (частково) навчання. Тому не рекомендується відключати масштабування, за винятком тих випадків, коли діапазон вихідних значень дуже малий і розташований поблизу нуля. Це ж розуміння свідчить на користь масштабування і при передобробці у перцептронах (при якій, у принципі, ваги першого проміжного шару можна легко коректувати, домагаючись цим будь-якого потрібного масштабування).

Радіальні мережі по своїй природі нездатні до екстраполяції. Чим далі вхідний приклад розташований від точок, що відповідають радіальним елементам, тим менше стають рівні активації радіальних елементів і (зрештою) тим менше буде

вихідний сигнал мережі. Вхідний приклад, розташований далеко від центрів радіальних елементів, дасть нульовий вихідний сигнал. Прагнення мережі не екстраполювати дані можна вважати достоїнством (це залежить від предметної області і думки користувача), однак убування вихідного сигналу (на перший погляд) не є достоїнством. Якщо ми прагнемо уникати екстраполяції, то для вхідних точок, що відрізняються великим ступенем новизни, у якості виходу ми, як правило, хочемо мати усереднене значення.

Для радіальних мереж у задачах регресії цього можна досягти за допомогою масштабуючої функції середнє/стандартне відхилення. Навчальні дані масштабуються таким чином, щоб середнє вихідне значення дорівнювало 0,0, а всі інші значення були б промасштабовані на стандартне відхилення вихідних сигналів. При обробці вхідних точок, що лежать поза областями дії радіальних елементів, вихідний сигнал мережі буде приблизно дорівнює середньому значенню.

*Якість роботи мережі в задачі регресії можна перевірити декількома способами.*

По-перше, мережі можна повідомити вихідне значення, що відповідає будь-якому спостереженню (чи якомусь новому спостереженню, що необхідно перевірити). Якщо це спостереження містилося у вихідних даних, то видається значення різниці (нев'язання).

По-друге, можуть бути отримані підсумкові статистики. До них відносяться середнє значення і стандартне відхилення, обчислені для навчальних даних і для помилки прогнозу. У загальному випадку середнє значення помилки прогнозу буде дуже близьким до нуля (зрештою, нульове середнє для помилки прогнозу можна одержати, попросту оцінивши середнє значення навчальних даних і зовсім не звертаючись до значень вхідних змінних). Найбільш важливим показником є стандартне відхилення помилки прогнозу. Якщо воно не виявиться істотно менше стандартного відхилення навчальних даних, це буде означати, що мережа працює не краще, ніж проста оцінка по середньому. Далі, у пакеті Statistica Neural Networks користувачу видається відношення стандартного відхилення помилки прогнозу до стандартного відхилення навчальних даних. Якщо

воно істотно менше одиниці (наприклад, нижче 0,1), то це говорить про гарну якість регресії. Це регресійне відношення (точніше, величину одиниця мінус це відношення) іноді називають часткою поясненої дисперсії моделі.

По-третє, можна вивести зображення поверхні відгуку. Насправді, зрозуміло, ця поверхня являє собою  $N+1$ -мірний об'єкт, де  $N$  – число вхідних елементів, а вимір, що залишився, відповідає висоті точки на поверхні. Зрозуміло, що безпосередньо візуально представити таку поверхню при  $N$  більшому двох неможливо (а реально  $N$  завжди більше двох). Проте, у пакеті Statistica Neural Networks можна виводити зрізи поверхні відгуку по будь-яким двом вхідним змінним. При цьому значення всіх інших вхідних змінних фіксуються, і змінюються тільки дві обрані. Всім іншим змінним можна додати будь-яке значення за своїм розсудом (за замовчуванням система Statistica Neural Networks візьме для них середні значення). Значення двох досліджуваних змінних можна змінювати в довільному діапазоні (за замовчуванням – у діапазоні зміни навчальних даних).

У задачах аналізу часових рядів метою є прогноз майбутніх значень змінної, залежної від часу, на основі попередніх значень її та/або інших змінних. Як правило, прогнозована змінна є числовою, тому прогнозування часових рядів – це окремий випадок регресії. Однак таке обмеження не закладене в пакет Statistica Neural Networks, так що в ньому можна прогнозувати і часові ряди номінальних змінних.

Звичайно чергове значення часового ряду прогнозується по деякому числу його попередніх значень (прогноз на один крок уперед у часі). У пакеті Statistica Neural Networks можна виконувати прогноз на будь-яке число кроків. Після того, як обчислене чергове передбачуване значення, воно підставляється назад і з його допомогою (а також попередніх значень) виходить наступний прогноз – це називається проекцією часового ряду. У пакеті Statistica Neural Networks можна здійснювати проекцію часового ряду і при покроковому прогнозуванні. Зрозуміло, що надійність такої проекції тим менше, чим більше кроків уперед ми намагаємося пророчити. У випадках, коли потрібно зовсім